

Maestría en Ciencias Actuariales

Estimación de la frecuencia y severidad asociada a pacientes recuperados del covid-19 en la costa caribe colombiana.

Juan Diego Torres Martinez

Bogotá, D.C., sábado 3 de junio de 2023

Estimación de la frecuencia y severidad asociada a pacientes recuperados del covid-19 en la costa caribe colombiana.

Tesis para optar al título de Magíster en Ciencias Actuariales

Ingrid Johana Quizacan Rubio

Director

Bogotá, D.C., sábado 3 de junio de 2023

La tesis de maestría titulada “Estimación de la frecuencia y severidad asociada a pacientes recuperados del covid-19 en la costa caribe colombiana”, presentada por Juan Diego Torres Martinez, cumple con los requisitos establecidos para optar al título de Magíster en Ciencias Actuariales.

Director de la tesis

Ingrid Johana Quizacan Rubio

Jurado 1

Yesid Esteban Clavijo Penagos

Jurado 2

Maria Carolina Ramirez Sepulveda

Bogotá, D.C., sábado 3 de junio de 2023

Dedicatoria

A mi familia, amigos y colegas quienes me brindaron su apoyo durante todo este proceso.

La pandemia generada por el covid-19 fue un evento el cual desafió la eficacia de la mayoría de los sistemas sanitarios del planeta, muchos de los cuales colapsaron de forma abrupta desencadenando un número de muertes el cual, en 2023 asciende a más de 6 millones de personas repartidas en más de 680 millones de casos alrededor del mundo¹, en Colombia, de acuerdo con cifras del instituto nacional de salud, se han presentado más de 6 millones de casos dentro de los cuales se reparten más de 142 mil muertes las cuales se concentran mayormente en adultos mayores de 60 años².

El confinamiento obligatorio decretado por el gobierno nacional colombiano en el año 2020 tuvo su fin el 30 de junio de 2022, en donde 83% de los colombianos contaba con al menos una dosis de la vacuna y más del 70% de la población contaba con esquema completo.

Sin embargo, existe un fenómeno que cada vez es más estudiado, corresponde al Covid prolongado o Long Covid. El cual consiste en la prevalencia de síntomas una vez superada la enfermedad.

Con el fin de comprender los efectos del Long Covid en materia de Frecuencia y Severidad para la toma de decisiones se ha escogido una muestra de la población recuperada del covid-19 en los departamentos de Atlántico, Bolívar, Córdoba, Magdalena, Sucre, y Cesar. De esta manera se procedió a realizar una serie de preguntas destinadas a corroborar la existencia de Síntomas PostCovid u otras situaciones que afecten de forma negativa la calidad de vida de los individuos.

Los resultados arrojaron que el 68,21% de los individuos ha manifestado por lo menos un síntoma PostCovid, y más de un 50% Manifiesta alguna situación que afecta su calidad de vida.

Lo anterior deja en evidencia que todas aquellas personas recuperadas del covid-19 poseen un riesgo en salud latente el cual puede ser mayor o menor para cada individuo.

Así mismo, se observó que las personas recuperadas del covid-19 tienen una frecuencia y severidad mayor respecto a cuando no estaban infectadas. Adicionalmente la frecuencia y severidad esperadas para el año 2023 es mayor en los afiliados recuperados de covid-19 en comparación con aquellos que nunca se infectaron

Lo anterior se pudo constatar puesto que las personas recuperadas del covid 19 poseen una frecuencia y severidad mayor respecto a cuando no estaban infectadas. Adicionalmente la frecuencia y severidad esperadas para el año 2023 es mayor en los afiliados recuperados de covid-19 en comparación con aquellos que nunca se infectaron.

¹ <https://www.trt.net.tr/espanol/covid19>

² <https://www.ins.gov.co/Noticias/paginas/coronavirus.aspx>

Índice general

Contenido

Índice general.....	5
Índice de tablas.....	7
Índice de figuras	9
Introducción	11
Capítulo I - Objetivos	13
Objetivo General.....	13
Objetivos Específicos.....	13
Capítulo II - Marco Teórico.	14
Conceptos actuariales en salud	14
COVID-19 y atención post-COVID	14
Siniestros en salud y gestión del riesgo en salud	16
Demanda de servicios de salud	17
Modelos para estimar el número de siniestros por usuario.....	18
Modelos para estimar la severidad de los siniestros.	21
Capítulo III - Metodología.....	23
Rastreo de síntomas Post-covid	24
Estimación del tamaño de la muestra de la población Objetivo.	25
Definición de periodos Pre, Intra, y Post covid.....	26
Número de utilizaciones por afiliado $u_{\tau\gamma}$	27
Ajuste del modelo	27
Selección del modelo	28
Costo por Usuario $X_{\tau\gamma}$	30
Ajuste del modelo	30
Selección del modelo	30
Capítulo III - Resultados.....	33
Análisis descriptivo – población objetivo.	33
Síntomas PostCovid y nivel de calidad de vida.....	36
Incremento en los costos de atención (PreCovid Vs PostCovid)	41

Incremento en el número de utilizaciones (PreCovid Vs PostCovid).....	43
Utilizaciones esperadas 2023.....	45
Ajuste de funciones de densidad de probabilidad – Afiliados no infectados por covid-19	45
Ajuste de funciones de densidad de probabilidad – Afiliados Recuperados del Covid-19	53
Resumen - Utilizaciones Esperadas 2023 (Usuarios Recuperados del Covid Vs No Infectados)	60
Costo Esperado por Afiliado 2023	62
Ajuste de funciones de densidad de probabilidad – Afiliados no infectados por covid-19	62
Ajuste de funciones de densidad de probabilidad – Afiliados Recuperados del Covid-19	70
Resumen - Costo Esperado por Afiliado 2023 (Usuarios Recuperados del Covid Vs No Infectados).....	79
Capítulo III – Conclusiones	80
Capítulo IV – Trabajos Futuros.	81
Referencias.....	82
Anexos.....	85
Anexo 1. Encuesta Aplicada.	85
Anexo 2. Código en R	86

Índice de tablas

Tabla 1. Estadísticas Globales de encuesta Post COVID.	36
Tabla 2. Distribución Regional de encuestas Post COVID realizadas.	36
Tabla 3. Distribución de respuestas afirmativas – Bloque de preguntas Post COVID.	38
Tabla 4. Distribución de respuestas afirmativas – Bloque de preguntas Calidad de Vida.	39
Tabla 5. Incremento de Costo medio de Atención PreCovid Vs PostCovid (2019 Vs 2022)	41
Tabla 7. Incremento del número de utilizaciones por usuario – Pre COVID Vs Post COVID (2019 Vs 2022)	43
Tabla 8. Incremento Promedio en las utilizaciones (Pre Covid Vs Post Covid) (2019 Vs 2022)	44
Tabla 11. Criterios de Ajuste – ui, t 1 Regional Atlántico.	45
Tabla 12. Parámetros del modelo seleccionado - Regional Atlántico.	46
Tabla 13. Criterios de Ajuste – ui, t 1 Regional Bolívar.	47
Tabla 14. Parámetros del modelo seleccionado - Regional Bolívar.	48
Tabla 15. Criterios de Ajuste – ui, t 1 Regional Córdoba.	48
Tabla 16. Parámetros del modelo seleccionado - Regional Córdoba.	49
Tabla 17. Criterios de Ajuste – ui, t 1 Regional Magdalena.	49
Tabla 18. Parámetros del modelo seleccionado - Regional Magdalena.	50
Tabla 19. Criterios de Ajuste – ui, t 1 Regional Sucre.	50
Tabla 20. Parámetros del modelo seleccionado - Regional Sucre.	51
Tabla 21. Criterios de Ajuste – ui, t 1 Regional Cesar.	52
Tabla 22. Parámetros del modelo seleccionado - Regional Bolívar.	52
Tabla 23. Criterios de Ajuste – ui, t 2 Regional Atlántico.	53
Tabla 24. Parámetros del modelo seleccionado - Regional Bolívar.	54
Tabla 25. Criterios de Ajuste – ui, t 2 Regional Bolívar.	54
Tabla 26. Parámetros del modelo seleccionado - Regional Bolívar.	55
Tabla 27. Criterios de Ajuste – ui, t 2 Regional Córdoba.	55
Tabla 28. Parámetros del modelo seleccionado - Regional Bolívar.	56
Tabla 29. Criterios de Ajuste – ui, t 2 Regional Magdalena.	57
Tabla 30. Parámetros del modelo seleccionado - Regional Magdalena.	58
Tabla 31. Criterios de Ajuste – ui, t 2 Regional Sucre.	58
Tabla 32. Parámetros del modelo seleccionado - Regional Sucre.	59
Tabla 33. Parámetros del modelo seleccionado - Regional Sucre.	60
Tabla 34. Incremento Utilizaciones Esperadas 2023.	61
Tabla 35. Criterios de Ajuste – xi, t 1 Regional Atlantico.	62
Tabla 36. Parámetros del modelo seleccionado - Regional Atlántico.	63
Tabla 37. Criterios de Ajuste – xi, t 1 Regional Bolívar.	63
Tabla 38. Parámetros del modelo seleccionado - Regional Bolívar.	64
Tabla 39. Criterios de Ajuste – xi, t 1 Regional Córdoba.	65
Tabla 40. Parámetros del modelo seleccionado - Regional Córdoba.	66
Tabla 41. Criterios de Ajuste – xi, t 1 Regional Magdalena.	66
Tabla 42. Criterios de Ajuste – xi, t 1 Regional Magdalena.	67
Tabla 43. Criterios de Ajuste – xi, t 1 Regional Sucre.	67
Tabla 44. Parámetros del modelo seleccionado - Regional Sucre.	68
Tabla 45. Criterios de Ajuste – xi, t 1 Regional Cesar.	69
Tabla 46. Parámetros del modelo seleccionado - Regional Cesar.	70

Tabla 47. Criterios de Ajuste – χ^2 , t Regional Atlántico.	70
Tabla 48. Parámetros del modelo seleccionado - Regional Atlántico.	71
Tabla 49. Criterios de Ajuste – χ^2 , t Regional Bolívar.	72
Tabla 50. Parámetros del modelo seleccionado - Regional Bolívar.	72
Tabla 51. Parámetros del modelo seleccionado - Regional Córdoba.	74
Tabla 52. Criterios de Ajuste – χ^2 , t Regional Magdalena.	74
Tabla 53. Parámetros del modelo seleccionado - Regional Magdalena.	75
Tabla 54. Criterios de Ajuste – χ^2 , t Regional Sucre.	76
Tabla 54. Parámetros del modelo seleccionado - Regional Sucre.	77
Tabla 55. Criterios de Ajuste – χ^2 , t Regional Cesar.	77
Tabla 56. Parámetros del modelo seleccionado - Regional Bolívar.	78
Tabla 57. Incremento Costo Esperado por Afiliado 2023.	79

Índice de figuras

Figura 1. Distribución Etaria de Recuperados Del covid-19.....	12
Figura 2. Diagrama de Pareto – Población recuperada del covid-19 por departamento.....	33
Figura 3. Pirámide poblacional – Población Recuperada del Covid-19.....	34
Figura 4. Pirámide poblacional de población recuperada del COVID 19 – Atlántico, Bolívar, y Córdoba.....	35
Figura 5. Pirámide poblacional de población recuperada del COVID 19 Cesar, Magdalena y Sucre.....	35
Figura 6. Comparativo - Presencia de síntomas Post COVID por régimen y regional.....	40
Figura 7. Comparativo – Dificultades en la calidad de vida por régimen y regional.	40
Figura 8. Incremento de Costo medio de Atención PreCovid Vs PostCovid (2019 Vs 2022) Por Regional.	42
Figura 9. Incremento de las utilizaciones promedio por usuario PreCovid Vs PostCovid (2019 Vs 2022) Por Regional. .	43
Figura 10. Gráficos Q-Q – ui, t 1 Regional Atlántico.	46
Figura 11. Gráficos Q-Q – ui, t 1 Regional Bolívar.	47
Figura 12. Gráficos Q-Q – ui, t 1 Regional Córdoba.....	48
Figura 13. Gráficos Q-Q – ui, t 1 Regional Magdalena.	50
Figura 14. Gráficos Q-Q – ui, t 1 Regional Sucre.	51
Figura 15. Gráficos Q-Q – ui, t 1 Regional Cesar.....	52
Figura 16. Gráficos Q-Q – ui, t 2 Regional Atlántico.....	53
Figura 17. Gráficos Q-Q – ui, t 2 Regional Bolívar.....	55
Figura 18. Gráficos Q-Q – ui, t 2 Regional Córdoba.....	56
Figura 19. Gráficos Q-Q – ui, t 2 Regional Magdalena.....	57
Figura 20. Gráficos Q-Q – ui, t 2 Regional Sucre.....	58
Figura 21. Criterios de Ajuste – ui, t 2 Regional Cesar.....	60
Figura 22. Gráficos Q-Q – xi, t 1 Regional Atlántico.....	62
Figura 23. Gráficos Q-Q – xi, t 1 Regional Bolívar.....	64
Figura 24. Gráficos Q-Q – xi, t 1 Regional Córdoba.....	65
Figura 25. Gráficos Q-Q – xi, t 1 Regional Magdalena.....	67
Figura 26. Gráficos Q-Q – xi, t 1 Regional Sucre.....	68
Figura 27. Gráficos Q-Q – xi, t 1 Regional Cesar.....	69
Figura 28. Gráficos Q-Q – xi, t 2 Regional Atlántico.....	71
Figura 29. Gráficos Q-Q – xi, t 2 Regional Bolívar.....	72
Figura 30. Gráficos Q-Q – xi, t 2 Regional Córdoba.....	73
Figura 31. Gráficos Q-Q – xi, t 2 Regional Magdalena.....	75
Figura 32. Gráficos Q-Q – xi, t 2 Regional Sucre.....	76
Figura 33. Gráficos Q-Q – xi, t 2 Regional Cesar.....	78

Introducción

La pandemia de COVID-19 es una crisis sanitaria global sin precedentes causada por el virus SARS-CoV-2, que se identificó por primera vez en Wuhan, China, en diciembre de 2019 (World Health Organization, 2020). La enfermedad, conocida como COVID-19, se propagó rápidamente a nivel mundial, afectando a millones de personas y causando un impacto significativo en la salud pública y los sistemas sanitarios.

El virus se propaga principalmente a través de partículas líquidas expulsadas al toser, estornudar o hablar. Los síntomas varían desde leves hasta graves, incluyendo fiebre, tos y dificultad para respirar. En casos graves, puede provocar neumonía, insuficiencia orgánica y, en última instancia, la muerte. Los grupos de mayor riesgo incluyen personas mayores y aquellos con afecciones médicas preexistentes, como enfermedades cardíacas, diabetes y enfermedades pulmonares.

Los efectos de la pandemia en la salud pública y los sistemas sanitarios han sido devastadores. La rápida propagación del virus y la falta de tratamientos y vacunas iniciales sobrecargaron a los hospitales y sistemas de atención médica en todo el mundo. Muchos países experimentaron escasez de suministros médicos esenciales, como ventiladores y equipo de protección personal (EPP)

Además, la pandemia también tuvo efectos indirectos en la salud pública, ya que los recursos médicos se desviaron hacia la lucha contra el COVID-19, lo que resultó en la interrupción de servicios médicos esenciales para otras enfermedades y condiciones. El impacto económico también contribuyó a un deterioro en la salud mental de muchas personas, con un aumento en casos de ansiedad y depresión. Un artículo publicado por organización panamericana de la salud (OPS) en junio de 2021 concluye que el 75% de los países adscritos a dicha organización confirmaron una grave afectación e interrupción los servicios y programas de prevención y promoción de salud mental, en los que los servicios de prevención del suicidio, servicios para niños y servicios para adultos mayores han sido los más afectados respectivamente (OPS, 2021).

En el contexto actual, la pandemia de COVID-19 sigue siendo relevante debido a la aparición de nuevas variantes del virus y la necesidad de mantener la vigilancia y el esfuerzo en la prevención y el control de la enfermedad. Lo anterior dado que Diversos estudios han comprobado la presencia de síntomas Post-covid mediante el uso de encuestas hacia los infectados, uno de estos estudios encontró que el 74% de los pacientes encuestados presentó dificultad respiratoria (Weerahandi, 2021), del mismo modo, Anaya et al presentan resultados en los que más de un 90% de la muestra de pacientes recuperados presentan al menos un síntoma. Así mismo un estudio presentado por la organización FairHealth en junio de 2021 el cual fue realizado sobre una población de 1,959,982 que fueron diagnosticados con covid concluyó que: De los pacientes que fueron hospitalizados por COVID-19, el porcentaje que tuvo una condición post-COVID fue de 50%, así mismo en aquellos pacientes sintomáticos, pero no hospitalizados el porcentaje fue del 27,5%, y en los pacientes que fueron asintomáticos el porcentaje fue del 19%.

Ahora bien, en el contexto local. Según Cifras del Instituto Nacional de Salud (INS) a corte del día 2 de junio de 2023 se han confirmado 6.368.934 casos positivos en el país, de los cuales 6.189.718 corresponden a pacientes recuperados.³ A continuación, se presenta la distribución de la población recuperada por grupo etario.

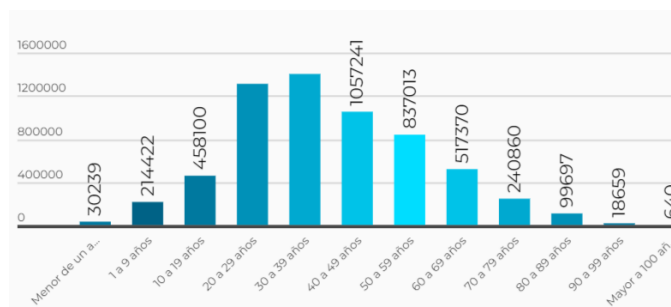


Figura 1. Distribución Etaria de Recuperados Del covid-19
Fuente. Instituto Nacional de Salud.

Tal y como se aprecia, la mayor parte de la población recuperada se concentra en las edades entre los 20 y los 60 años. Lo anterior es de particular interés debido a que, tal y como se mencionó anteriormente todas aquellas personas recuperadas se encuentran expuestas a manifestar sintomatología PostCovid.

Una vez se ha mencionado lo anterior, desde el punto de vista del sector asegurador en salud es sumamente importante tener en cuenta los efectos del Long Covid puesto que este se convierte en un riesgo en salud adicional el cual puede tener efectos a largo plazo en la salud de las personas, resultando así en una mayor demanda de servicios en salud, lo que a su vez incrementaría los costos de atención.

³ <https://www.ins.gov.co/Noticias/Paginas/coronavirus.aspx>

Por ende, el presente estudio se basa en la caracterización de la población recuperada del COVID-19, a través del análisis de variables como el sexo, la edad y la ubicación geográfica. Posteriormente, se lleva a cabo un análisis de los resultados obtenidos mediante la implementación de una encuesta diseñada para identificar los síntomas post-COVID y las condiciones que afectan la calidad de vida de las personas recuperadas de este.

Asimismo, se realiza un análisis comparativo de la frecuencia y severidad antes y después del COVID-19. Con base en estos hallazgos, se procede a estimar la frecuencia y severidad esperadas, con el propósito de contrastar el incremento previsto para el próximo año en relación con aquellos individuos no infectados por el SARS-CoV-2.

Este estudio se enfoca en abordar de manera rigurosa la comprensión de las secuelas del COVID-19 en la población recuperada, con el objetivo de aportar evidencia sólida y fundamentada que contribuya a la toma de decisiones informadas y al diseño de estrategias adecuadas para la atención de las personas afectadas.

Capítulo I - Objetivos

Para dar solución a la problemática planteada se definen los siguientes objetivos, los cuales establecen el propósito central del presente estudio.

Objetivo General

Estimar el costo y el número futuro de siniestros por usuario para la población recuperada del covid-19 a partir de la caracterización y el análisis de la demanda de servicios de salud para la toma de decisiones asociadas la planeación y despliegue de estrategias encaminadas a la mitigación del riesgo en salud de la población objetivo.

Objetivos Específicos

- Caracterizar a la población afiliada que se ha recuperado del covid-19 mediante el uso de variables sociodemográficas, epidemiológicas, y su demanda histórica de servicios en salud para la realización de un análisis descriptivo el cual permita identificar patrones y tendencias en la población objetivo.
- Segmentar a la Población de interés mediante la agrupación de los afiliados con características comunes tal que esto permita un correcto proceso de modelado de la frecuencia y severidad de los grupos.
- Evaluar el ajuste de los modelos aplicados para la estimación de la frecuencia y severidad de cada uno de los segmentos mediante un análisis comparativo el cual permita seleccionar el modelo idóneo para cada uno de estos.

Capítulo II - Marco Teórico.

En este capítulo se abordan los conceptos clave relacionados con la atención post-COVID, siniestros en salud, gestión del riesgo en salud, demanda de servicios de salud, y segmentación de poblaciones. Además, se analizan las teorías y modelos existentes en la literatura científica para proporcionar un fundamento sólido para la investigación.

Conceptos actuariales en salud

Esta sección del marco teórico se centra en los conceptos clave relacionados con la actuaria en salud, proporcionando una base sólida para el análisis de la siniestralidad y el riesgo en salud asociado a la atención post-COVID.

- *Definición de siniestralidad y riesgo en salud:* La siniestralidad se refiere al número y costo de los siniestros o reclamaciones presentadas por los asegurados durante un período determinado (Charpentier, 2005). El riesgo en salud, por otro lado, se refiere a la probabilidad de ocurrencia de un evento no deseado, evitable y negativo para la salud del individuo, que puede ser también el empeoramiento de una condición previa, la necesidad de requerir más consumo de bienes y servicios que hubiera podido evitarse. El evento, la ocurrencia de la enfermedad / traumatismos o su evolución desfavorable y sus causas son los diferentes factores asociados (MinSalud, 2018).
- *Modelos actuariales para la estimación de costos y siniestros en salud:* Los modelos actuariales en salud son herramientas matemáticas y estadísticas utilizadas para evaluar y predecir los costos y siniestros en salud, tomando en cuenta factores como la frecuencia, la severidad y la duración de los eventos adversos para la salud (Boland et al., 2018). Estos modelos incluyen enfoques determinísticos y estocásticos, como los modelos de regresión, las tablas de vida y los modelos de supervivencia, que permiten a los actuarios y a los responsables de la gestión del riesgo en salud hacer estimaciones precisas y fundamentadas sobre la siniestralidad futura (Garrido et al., 2014).

COVID-19 y atención post-COVID

La enfermedad por coronavirus (COVID-19) es una enfermedad infecciosa causada por el virus SARS-CoV-2. La mayoría de las personas infectadas con el virus experimentarán una enfermedad respiratoria de leve a moderada y se recuperarán sin requerir un tratamiento especial. Sin embargo, algunos se enfermarán gravemente y requerirán atención médica. Las personas mayores y aquellas con afecciones médicas subyacentes, como enfermedades cardiovasculares, diabetes, enfermedades respiratorias crónicas o cáncer, tienen más probabilidades de desarrollar enfermedades graves. Cualquiera puede enfermarse de COVID-19 y enfermarse gravemente o morir a cualquier edad. (World Health Organization, 2023)

El tiempo medio desde el inicio de los síntomas hasta la recuperación es de 2 semanas cuando la enfermedad ha sido leve y 3-6 semanas cuando ha sido grave o crítica. (Ministerio de Sanidad de España, 2021)

La atención post-COVID, también conocida como síndrome post-agudo secuelar del SARS-CoV-2 (PASC, por sus siglas en inglés), se refiere al tratamiento y seguimiento médico de los individuos que han superado la fase aguda del COVID-19 pero continúan experimentando síntomas persistentes o desarrollan secuelas a largo plazo después de la infección inicial (Munipalli, 2022).

La atención post-COVID ha emergido como un área de creciente preocupación en la gestión de la salud pública, en todo el mundo, con el creciente número de sobrevivientes de COVID-19, se ha prestado cada vez más atención se han sentido atraídos por las secuelas prolongadas o tardías de la infección por SARS-CoV-2, que se conocen coloquialmente como síndrome de "COVID prolongado" o "Long Covid" (Stefanou MI, 2022).

La literatura científica sobre la atención post-COVID sigue en expansión, abordando aspectos como la prevalencia y duración de los síntomas, los factores de riesgo, los mecanismos biológicos subyacentes y las intervenciones terapéuticas, por lo tanto, existen múltiples definiciones acerca de las secuelas presentadas en pacientes recuperados del covid-19.

M-I Stefanou, et al. Presentan una tabla que consolida las distintas nomenclaturas y sus respectivas descripciones.

Nomenclatura	Descripción
Long COVID	Síntomas de COVID-19 a largo plazo caracterizados por un curso cíclico, progresivo o multifásico. (Callard And Perego, 2021)
Long COVID	Síntomas que persisten durante mas 2 meses. (Brodin P, 2021)
Long COVID	Signos y síntomas que persisten por más de 4 semanas y pueden atribuirse a la infección por COVID-19. (Sivan M and Taylor S, 2020)
Long-haul COVID Long-tail COVID	Síntomas que persisten durante mas 100 días. (Nath A, 2020)
Long post-COVID symptoms	Síntomas que duran de 12 a 24 semanas. (Fernández de Las Peñas C, 2021)
Post-acute COVID syndrome (PACS)	Síntomas persistentes y/o tardíos o a largo plazo complicaciones más allá de las 4 semanas desde el inicio de los síntomas. (Nalbandian A, 2021)

Post-acute sequelae of SARS-CoV-2 (PASC)	Síntomas que persisten durante mas 1 mes. (Oh ES, 2021)
Post-COVID-19	Síntomas que persisten durante mas 2 meses. (Davido B, 2020)
Persistent post-COVID symptoms	Síntomas que persisten durante mas 24 semanas. (Fernández de Las Peñas C, 2021)
Post-COVID-19 syndrome	Signos y síntomas que se desarrollan durante o después de una Infección compatible con COVID-19, presente durante >12 semanas y no son atribuibles a diagnósticos alternativos (Shah W, 2021)

Siniestros en salud y gestión del riesgo en salud

Los siniestros en salud se refieren a eventos adversos o desfavorables asociados con la atención médica que pueden resultar en consecuencias negativas para la salud de los pacientes y en costos adicionales para los sistemas sanitarios. Estos eventos pueden incluir hospitalizaciones prolongadas, intervenciones quirúrgicas, tratamientos médicos costosos, enfermedades crónicas, discapacidades y pérdidas económicas, tanto para los pacientes como para los proveedores de atención médica (Thomas Golaszewski, 1989)

La Gestión Integral del Riesgo en Salud Pública se entiende como el conjunto de procesos y acciones de carácter técnico, político, económico, social y cultural que permiten la gestión de las circunstancias de vida que representan amenazas individuales, familiares y colectivas para el goce efectivo del más alto nivel de salud 105 posible, mediante la identificación, caracterización, medición, intervención, seguimiento, monitoreo y evaluación de las estrategias promocionales de la salud y de los riesgos/amenazas generadoras de morbilidad, discapacidad, mortalidad evitable y desigualdad; cuyas respuestas requieren de la articulación y coordinación de los activos sociales en salud para su resolución (MinSaud, 2022)

Demanda de servicios de salud

La demanda de servicios de salud hace referencia a la necesidad de atención médica por parte de una población, la cual puede manifestarse a través de consultas médicas, procedimientos quirúrgicos, tratamientos farmacológicos y otros servicios relacionados con la promoción, prevención, diagnóstico, tratamiento y rehabilitación de la salud (Folland, 2017)

Varios factores influyen en la demanda de servicios de salud, entre los que se incluyen:

- Factores demográficos: La edad, el género, la etnia y la estructura familiar pueden afectar la demanda de servicios de salud, ya que diferentes grupos poblacionales pueden tener diferentes necesidades y patrones de utilización de atención médica (Andersen, 1995)
- Factores socioeconómicos: El nivel de ingresos, la educación, el empleo y el seguro de salud también pueden influir en la demanda de servicios de salud, ya que estos factores pueden afectar la capacidad de las personas para acceder y pagar la atención médica (Gerdtham and Jönsson, 2000)
- Factores de salud: La prevalencia de enfermedades crónicas, la carga de enfermedades infecciosas y las condiciones de salud mental en una población pueden influir en la demanda de servicios de salud, ya que las personas con condiciones de salud específicas pueden requerir diferentes tipos de atención médica (Wagstaff & Claeson, 2004)

En el contexto de la pandemia de COVID-19 y la atención post-COVID, la demanda de servicios de salud ha experimentado cambios significativos. La aparición de la enfermedad y sus secuelas a largo plazo en muchos pacientes recuperados ha generado un aumento en la demanda de atención médica, incluidas consultas de seguimiento, tratamientos de rehabilitación, pruebas diagnósticas y manejo de síntomas persistentes (Ziyad Al-Aly, 2021).

Modelos para estimar el número de siniestros por usuario.

La estimación de la función de densidad de probabilidad se realiza por medio de modelos paramétricos. De acuerdo con (Mikosch T. , 2009) algunas de las funciones de densidad de probabilidad discretas que usualmente se usan para modelar la frecuencia son: Poisson, y Geométrica las cuales serán incluidas en este estudio. En complemento, se utilizarán otras distribuciones para variables discretas como lo son la logarítmica Binomial Negativa Tipo I y Poisson Generalizadas definidas por Rigby et al. En su libro “*Distributions for Modeling Location, Scale, and Shape Using GAMLSS in R.*”

Distribución de Poisson.

La distribución de Poisson se utiliza para modelar la ocurrencia de eventos raros en un intervalo de tiempo o espacio específico. Este describe la probabilidad de que ocurra un número determinado de eventos en un período dado, cuando los eventos ocurren de manera independiente y a una tasa promedio constante. (Mikosch T. , 2009)

La función de probabilidad de la distribución de Poisson se define como:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad (1)$$

donde (X) es el número de siniestros, (k) es un número entero no negativo, (λ) es la tasa promedio de siniestros y (e) es la base del logaritmo natural.

Mixturas explícitas de distribuciones de Poisson:

Suponga que dado una variable aleatoria γ toma el valor γ , Y tiene una distribución de Poisson con media $\mu\gamma$, es decir, $Y|\gamma \sim PO(\mu\gamma)$ donde $\mu > 0$, así mismo se supone que γ tiene una función de densidad de probabilidad $f_\gamma(\gamma)$ definida en $(0, \infty)$, entonces la distribución marginal de Y es una mixtura de la distribución de poisson.

Por ejemplo, suponga que $Y|\gamma \sim PO(\mu\gamma)$ donde $\gamma \sim GA\left(1, \sigma^{\frac{1}{2}}\right)$. La distribución marginal es $Y \sim NBI(\mu, \sigma)$.

Para mostrar este resultado se tiene que:

$$P(Y = y|\gamma) = \frac{e^{-\mu\gamma} (\mu\gamma)^y}{y!}$$

Donde:

$$f_\gamma(\gamma) = \frac{\gamma^{\frac{1}{\sigma}-1} \exp\left(-\frac{\gamma}{\sigma}\right)}{\sigma^{\left(\frac{1}{\sigma}\right)} \Gamma\left(\frac{1}{\sigma}\right)}, \gamma > 0$$

Por tanto, la ecuación queda de la siguiente manera:

$$\begin{aligned}
 P(Y = \gamma) &= \int_0^{\infty} \frac{e^{-\mu\gamma} (\mu\gamma)^y}{y!} \cdot \frac{\gamma^{\frac{1}{\sigma}-1} \exp\left(-\frac{\gamma}{\sigma}\right)}{\sigma^{\left(\frac{1}{\sigma}\right)} \Gamma\left(\frac{1}{\sigma}\right)} d\gamma \\
 &= \frac{\Gamma\left(y + \frac{1}{\sigma}\right)}{\Gamma\left(\frac{1}{\sigma}\right) \Gamma(y + 1)} \left(\frac{\sigma\mu}{1 + \sigma\mu}\right)^y \left(\frac{1}{1 + \sigma\mu}\right)^{\frac{1}{\sigma}}
 \end{aligned} \tag{2}$$

Para $y = 0, 1, 2, 3, \dots$, donde $\mu > 0$ y $\sigma > 0$

Lo anterior corresponde a la función de densidad de probabilidad de la binomial negativa de tipo 1, $NBI(\mu, \sigma)$.

Distribución Binomial Negativa Tipo 1.

El modelo binomial negativo de tipo 1 denotado como $NBI(\mu, \sigma)$, Distribución mixta obtenida como la distribución marginal de Y cuando $Y|\gamma \sim POISSON(\mu\gamma)$ donde $\gamma \sim GAMMA\left(1, \sigma^{\frac{1}{2}}\right)$, es decir, γ posee una distribución gamma con media 1 y parámetro de escala $\sigma^{\frac{1}{2}}$.

La función de distribución de probabilidad se encuentra denotada por $NBI(\mu, \sigma)$, es:

$P(Y = y|\mu, \sigma)$ de acuerdo con la ecuación (2)

Distribución Geométrica.

La distribución geométrica es un tipo de distribución de probabilidad discreta que modela el número de ensayos independientes y repetidos necesarios hasta que ocurre el primer éxito. Se utiliza para calcular la probabilidad de que ocurra un evento por primera vez en un número determinado de ensayos. (George Casella, 2002)

Su función de probabilidad se encuentra dada por:

$$P(Y = y|\mu) = \frac{\mu^y}{(\mu + 1)^{y+1}}, \text{ para } x = 0, 1, 2, \dots, \text{ donde } \mu > 0 \tag{3}$$

Distribución Logarítmica.

Su función de probabilidad se encuentra dada por:

$$P(Y = y|\mu) = \frac{\alpha\mu}{y} \quad (4)$$

para $y = 1, 2, 3, \dots$, donde $\alpha = -[\log(1 - \mu)]^{-1}$ para $0 < \mu < 1$

Distribución Poisson Generalizada.

La distribución de Poisson generalizada es una aproximación de la distribución de binomial negativa generalizada. (Robert Rigby, 2017)

Su función de probabilidad se encuentra dada por:

$$P(Y = y|\mu, \sigma) = \left(\frac{\mu}{1 + \sigma\mu}\right)^y \frac{(1 + \sigma y)^{y-1}}{y!} \exp\left(\frac{-\mu(1 + \sigma y)}{1 + \sigma\mu}\right) \quad (5)$$

para $y = 0, 1, 2, \dots$ donde $\mu > 0$ y $\sigma > 0$

Modelos para estimar la severidad de los siniestros.

La estimación de la función de densidad de probabilidad se realiza por medio de modelos paramétricos. De acuerdo con (Mikosch T. , 2009) algunas de las funciones de densidad de probabilidad continuas que usualmente se usan para modelar la severidad son: log-normal, Weibull, exponencial Pareto y gamma, las cuales serán incluidas en este estudio. En complemento, se utilizarán otras distribuciones para variables continuas como lo son la distribución Burr XII y Gaussiana Inversa.

Se consideran las siguientes funciones de densidad de probabilidad:

Distribución Burr XII.

La distribución Burr XII, también conocida como distribución Singh-Maddala, es una distribución de probabilidad continua que se utiliza para modelar eventos con colas pesadas o colas largas. Esta distribución es un caso especial de la distribución Beta Generalizada Tipo II. (Kleiber C and Kotz S, 2003)

De acuerdo con (Robert Rigby, 2017). Su función de densidad de probabilidad está dada por:

$$f_Y(y|\mu, \sigma, \tau) = \frac{\sigma \tau \left(\frac{y}{\mu}\right)^\sigma}{y \left[1 + \left(\frac{y}{\mu}\right)^\sigma\right]^{\tau+1}} \quad (6)$$

para $y > 0$, donde $\mu > 0, \sigma > 0, \tau > 0$

Distribución Exponencial.

La distribución exponencial es una distribución de probabilidad continua la cual se utiliza a menudo para describir el tiempo de llegada de un suceso, como la avería de una máquina. (Mikosch T. , 2009)

De acuerdo con (Robert Rigby, 2017). Su función de densidad de probabilidad está dada por:

$$f_Y(y|\mu) = \frac{1}{\mu} \exp\left(-\frac{y}{\mu}\right) \text{ para } y > 0 \text{ donde } \mu > 0 \quad (1)$$

Distribución Gamma.

Se le conoce, también, como una generalización de la distribución exponencial. Es una distribución de probabilidad continua adecuada para modelizar el comportamiento de variables aleatorias con asimetría positiva y/o los experimentos en donde está involucrado el tiempo. (Arroyo, 2014)

De acuerdo con (Robert Rigby, 2017). Su función de densidad de probabilidad está dada por:

$$f_Y(y|\mu, \sigma) = \frac{y^{\left(\frac{1}{\sigma^2}-1\right)} e^{\left(\frac{-y}{\sigma^2\mu}\right)}}{(\sigma^2\mu)^{\frac{1}{\sigma^2}} \Gamma\left(\frac{1}{\sigma^2}\right)} \text{ para } y > 0, \quad \text{donde } \mu > 0, y \sigma > 0 \quad (7)$$

Distribución Gaussiana Inversa.

La distribución inversa gaussiana, también conocida como distribución de Wald, es una distribución de probabilidad continua que se utiliza para modelar variables positivas y asimétricas. (Klugman, 2019)

De acuerdo con (Robert Rigby, 2017). Su función de densidad de probabilidad está dada por:

$$f_Y(y|\mu, \sigma) = \frac{1}{\sqrt{2\mu\pi\sigma^2 y^3}} \left[-\frac{1}{2\mu^2\sigma^2 y} \cdot (y - \mu)^2 \right] \quad (8)$$

para $y > 0$, donde $\mu > 0, y \sigma > 0$

Distribución Log-normal.

La distribución logarítmico-normal, o Log-Normal, es una distribución de probabilidad continua que se utiliza para modelar variables que son el resultado de una transformación logarítmica de una variable normal. (Klugman, 2019)

De acuerdo con (Robert Rigby, 2017). Su función de densidad de probabilidad está dada por:

$$f_Y(y|\mu, \sigma) = \frac{1}{\sqrt{2\mu\pi\sigma^2}} \frac{1}{y} \exp\left\{-\frac{[\log(y) - \mu]^2}{2\sigma^2}\right\} \quad (9)$$

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

Distribución Pareto de un parámetro.

La distribución de Pareto se utiliza para modelar fenómenos con colas pesadas, lo que significa que existen eventos extremadamente raros, pero con impacto significativo. (McDonald, 1984)

De acuerdo con (Robert Rigby, 2017). Su función de densidad de probabilidad está dada por:

$$f_Y(y|\mu, \sigma) = \frac{\sigma \mu^\sigma}{y^{\sigma+1}}, \text{ para } y \geq 0, \mu > 0, y \sigma > 0 \quad (10)$$

Distribución Weibull.

La distribución Weibull es una distribución de probabilidad continua que se utiliza para modelar eventos con diferentes tasas de falla a lo largo del tiempo. (Cohen, 1973).

De acuerdo con (Robert Rigby, 2017). Su función de densidad de probabilidad está dada por

$$f_Y(y|\mu, \sigma) = \sigma \frac{y^{\sigma-1}}{\mu^\sigma} \exp \left[- \left(\frac{y}{\mu} \right)^\sigma \right], \text{ para } y > 0, \mu > 0, y \sigma > 0 \quad (11)$$

Capítulo III - Metodología

La selección población objetivo se hizo a partir de los casos diagnosticados con COVID-19 desde marzo de 2020 hasta el 31 de diciembre de 2022. Se eligió este lapso para tener una muestra lo suficientemente amplia y representativa de la población afectada por la enfermedad en un periodo prolongado. Además, se seleccionó información relevante sobre los pacientes como la fecha del primer diagnóstico, el departamento de atención, el género y la edad. Estos datos son esenciales para poder entender mejor la evolución y la distribución de la enfermedad en el territorio sobre el cual se circunscribe el presente estudio. Dicha población se encuentra repartida en 6 departamentos o Regionales. (Atlántico, Bolívar, Córdoba, Magdalena, Sucre, y Cesar).

Rastreo de síntomas Post-covid

Con el propósito de profundizar en el análisis de la población que ha logrado recuperarse de la COVID-19, se diseñó y aplicó una encuesta enfocada en los individuos que forman parte de este grupo demográfico. Esta encuesta se estructuró de manera que abordara diversos aspectos clave para obtener una perspectiva más amplia y detallada de las condiciones y circunstancias que enfrentan estos individuos. La encuesta completa puede encontrarse en el Anexo 1.

En primer lugar, se incluyeron preguntas relacionadas con la información sociodemográfica, como la edad, género, ocupación, nivel educativo y situación económica, entre otros, para poder contextualizar y comprender mejor las diferencias y similitudes entre los participantes.

En segundo lugar, se indagó acerca de los síntomas post-COVID, es decir, aquellos síntomas persistentes que puedan continuar afectando a los individuos incluso después de haber superado la fase aguda de la enfermedad.

Finalmente, para evaluar la calidad de vida de los participantes, se utilizaron preguntas extraídas del EQ-5D-5L (Cuestionario Europeo de 5 Dimensiones y 5 Niveles). Este instrumento, ampliamente reconocido y utilizado en investigaciones relacionadas con la salud, aborda temas cruciales como la movilidad, el autocuidado, la capacidad de realizar actividades cotidianas, la presencia de dolor o malestar y la ansiedad o depresión. La inclusión de este cuestionario permitió obtener una visión más holística de cómo la COVID-19 y sus secuelas han afectado el bienestar y la calidad de vida de la población recuperada.

Estimación del tamaño de la muestra de la población Objetivo.

Para este estudio destinado a evaluar a los pacientes recuperados de COVID-19, se empleó una metodología basada en el cálculo del tamaño de la muestra utilizando principios estadísticos estándar. Este cálculo asegura que la muestra seleccionada sea suficientemente grande para proporcionar resultados representativos y confiables.

El tamaño de la muestra fue calculado utilizando la fórmula de (Cochran, 1977), que es comúnmente utilizada en estadística para la determinación de muestras aleatorias simples. La fórmula de Cochran es la siguiente:

$$n = \frac{Z^2 p (1 - p)}{e^2} \quad (12)$$

Donde:

- n es el tamaño de la muestra
- Z es el valor Z para el nivel de confianza seleccionado. En este estudio, se ha seleccionado un nivel de confianza del 97%, por lo que el valor Z correspondiente es 2.17009.
- p es la proporción esperada del resultado en la población. Dado que no se cuenta con un valor específico para p , se ha tomado el valor que maximiza el tamaño de la muestra, que es 0.5.
- e es el error tolerado, en este caso, se ha elegido un error del 3% (0.03).

Adicionalmente, cuando la población es finita, es necesario ajustar el tamaño de la muestra utilizando la corrección de Cochran. La fórmula para la corrección de Cochran es la siguiente:

$$n_{\text{adj}} = \frac{n}{1 + \frac{(n - 1)}{N}} \quad (13)$$

Siendo N el tamaño de la población. En este estudio, se cuenta con una población de 111.446 pacientes recuperados de COVID-19.

Una vez aplicadas las ecuaciones (12) y (13) se obtiene que el tamaño de muestra necesario corresponde a 1.293 pacientes.

Definición de periodos Pre, Intra, y Post covid.

Para este apartado se toma como eje de la duración media de los síntomas brindada por el ministerio de sanidad de España mencionado en el apartado **COVID-19 y atención post-COVID**. Se selecciona el umbral máximo de 6 semanas para definir el periodo en el cual una persona ha superado la enfermedad.

Se establecen los siguientes criterios para definir los periodos en relación con el diagnóstico positivo para covid-19 de un afiliado i :

- **Periodo Pre-Covid:** Sea d_i la fecha del primer diagnóstico positivo para covid-19 del afiliado i . El periodo Pre-covid se define como el conjunto de siniestros S_{pre} tales que, para todo s en S_{pre} , se cumple que la fecha de ocurrencia f_s es estrictamente menor que d_i . En notación matemática:

$$S_{pre} = \{s | f_s < d_i\} \quad (14)$$

- **Periodo Intra-Covid:** Sea d_i la fecha del primer diagnóstico positivo para covid-19 del afiliado i . El periodo intra-covid se define como el conjunto de siniestros S tales que, para todo s en S , se cumple que la fecha de ocurrencia f_s está entre d_i y $d_i + 6$ *semanas*. En notación matemática:

$$S_{intra} = \{s | d_i \leq f_s \leq d_i + 6 \text{ semanas}\} \quad (15)$$

- **Periodo Post-Covid:** Sea d_i la fecha del primer diagnóstico positivo para covid-19 del afiliado i . El periodo Post-covid se define como el conjunto de siniestros S tales que, para todo s en S , se cumple que la fecha de ocurrencia f_s es estrictamente mayor que $d_i + 6$ *semanas*. En notación matemática:

$$S_{post} = \{s | f_s > d_i + 6 \text{ semanas}\} \quad (16)$$

A continuación, se introduce la función a trozos $\gamma_i(t)$ la cual tiene como finalidad definir la clasificación de cada siniestro en relación con el primer diagnóstico positivo para covid-19 del i – *esimo* afiliado, y se encuentra dada por:

$$\gamma_i(t) = \begin{cases} 1, & \text{si } t < d_i \\ 2, & \text{si } d_i \leq t \leq d_i + 6 \text{ semanas} \\ 3, & \text{si } t > d_i + 6 \text{ semanas} \end{cases} \quad (17)$$

Esta función se construye con base a las ecuaciones (14), (15) y (16) permitiendo la clasificación de los siniestros en tres periodos: Pre-Covid, Intra-Covid y Post-Covid, de

acuerdo con la relación temporal con el primer diagnóstico positivo para covid-19 del afiliado.

La función toma el valor 1 si la fecha de siniestro t es anterior a la fecha del primer diagnóstico positivo d_i del afiliado correspondiente. La función toma el valor 2 si t está entre d_i y $d_i + 6$ semanas, y toma el valor 3 si t es mayor que $d_i + 6$ semanas.

Número de utilizaciones por afiliado u_t^γ

Sea u_t^γ el número de las atenciones de cada usuario que utilizará las tecnologías del conjunto Δ en el año τ con clasificación γ siendo $\gamma = 1$ para afiliados que nunca fueron infectados por covid-19 y $\gamma = 2$ para para afiliados que se recuperaron por covid-19.

Se estima el valor esperado de u_t^γ , asumiendo que el comportamiento de la variable está gobernado por un modelo probabilístico que captura la incertidumbre sobre su verdadero valor. Es decir que u_t^γ sigue una distribución de probabilidad discreta con función de densidad $f(u; \theta)$, donde $u > 0$ y θ es un vector de parámetros desconocidos

Ajuste del modelo

Si las utilizaciones por usuario observadas de los periodos de interés, u_1, \dots, u_n , es una muestra de observaciones independientes de U_t , entonces se ajusta un modelo paramétrico conocido, $f(u; \theta)$, que maximiza la verosimilitud de la muestra de datos.

$$L(\theta) = \prod_{i=1}^n f(u_i; \theta), \quad (18)$$

en función de los parámetros del modelo. La expresión anterior se denomina función de verosimilitud y, aplicando logaritmo natural en ambos lados, se obtiene la función de log-verosimilitud:

$$l(\theta) = \sum_{i=1}^n \ln(f(u_i; \theta)). \quad (19)$$

Así pues, interesan los valores del vector de parámetros θ que maximizan el valor de la función $l(\theta)$, esto es:

$$\hat{\theta} = \operatorname{argmax}\{l(\theta)\}, \quad \theta \in \mathbb{R}^n \quad (20)$$

donde $\hat{\theta}$ es el vector de estimadores de máxima verosimilitud de θ . Finalmente, el modelo ajustado a los datos observados de las utilizaciones por usuario es la distribución de probabilidad con función de densidad $f(u; \hat{\theta})$ y vector de parámetros estimados $\hat{\theta}$.

Selección del modelo

En la ciencia estadística existen diferentes distribuciones de probabilidad discreta que se podrían ajustar a los datos observados de las utilizaciones por usuario. Suponga que existen n modelos rivales, con funciones de probabilidad, $f_1(u; \theta_1)$, $f_2(u; \theta_2)$, ..., $f_n(u; \theta_n)$, para $u > 0$ y θ_i es el vector de parámetros de la i -ésima función de probabilidad. La idea es replicar el procedimiento de maximización de la verosimilitud de los datos, estimación de parámetros y ajuste de cada uno de los modelos rivales, con el fin de seleccionar el mejor entre ellos según el grado de ajuste a los datos.

Estadísticos de bondad de ajuste Basados en la verosimilitud.

Sea $\{X_n^r\}_{n < N}$ una muestra de tamaño N_r de la variable aleatoria que representa los valores de las utilizaciones por afiliado para la regional r , con función de densidad de probabilidad $f_{x,r}(x)$. La función de verosimilitud de la muestra observada para cada regional, $L_r(\theta_r)$, está dada por la fórmula:

$$L_r(\theta_r) = \prod_{i=1}^{N_r} f_{x,r}(x_r; \theta_r), \quad (21)$$

donde θ_r es el vector de parámetros de la función de densidad de probabilidad para la regional r .

Ahora, si p es el número de parámetros de cada modelo para la regional r , los criterios de bondad de ajuste para la selección del mejor modelo son los siguientes:

- Criterio de Akaike: ofrece una estimación relativa de la información perdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos.

$$AIC = -2 \ln(L_r(\theta_r)) + 2p. \quad (22)$$

- Criterio de Akaike corregido: El AICC penaliza los modelos más complejos que tienen un mayor número de parámetros, evitando así el sobreajuste. Se prefiere un modelo con un valor de AICC más bajo, lo que indica un mejor equilibrio entre el ajuste del modelo y la complejidad.

$$AICC = -2 \ln(L_r(\theta_r)) + \frac{2N_r p}{N_r - p - 1}. \quad (23)$$

- Criterio de información bayesiano: es una medida de bondad de ajuste de un modelo estadístico, y es utilizado como un criterio para la selección de modelos entre un conjunto finito de modelos. Un BIC bajo implica un número menor de variables explicativas, mejor ajuste, o ambos.

$$BIC = -2 \ln(L_r(\theta_r)) + p \ln(N_r). \quad (24)$$

El mejor modelo es aquel con el menor valor del estadístico entre todos los criterios.

Estadísticos basados en la función de distribución empírica.

Esta clase de estadísticas se basa en la diferencia entre la estimación de la función de distribución acumulada (CDF) y la estimación de la función de distribución empírica (EDF). Sea $F_n(x)$ la función de distribución empírica y sea $F(x)$ la función de distribución acumulada.

- Test de Kolmogorov-Smirnov: se define como la distancia vertical máxima entre las funciones de distribución acumulada empíricas de dos muestras, o entre una función de distribución empírica y una función de distribución acumulada teórica de referencia. La ventaja principal de este estadístico es que es sensible a diferencias tanto en la localización como en la forma de la función de distribución acumulada:

$$KS = \max_{t \leq x \leq u} |F_n(x) - F(x)|, \quad (2)$$

donde t es un punto de truncamiento a izquierda y u es un punto de censura a la derecha.

- Estadístico de Anderson Darling: es un promedio ponderado de las diferencias al cuadrado entre la función empírica $F_n(x)$ y la función de distribución estimada $F(x)$:

$$AD = N \int_t^u \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]} f(x) dx. \quad (25)$$

Este estadístico tiende a poner más énfasis en el buen ajuste en las colas que en el medio de la distribución.

- Cramer von Mises: es un estadístico que es proporcional al valor esperado de la diferencia al cuadrado entre la función de distribución empírica y la función de distribución estimada.

$$CVM = N \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x). \quad (26)$$

Los criterios de selección consisten en ordenar y puntuar los n modelos rivales de acuerdo estadísticos bondad de ajuste basados en la verosimilitud, así como en aquellos basados en la distribución empírica. El mejor modelo es el que de acuerdo con los criterios anteriores tenga el mejor ajuste a los datos.

Una vez que se ranquean las distribuciones según el resultado de los criterios de información, con ayuda de los gráficos de contraste del Grafico Q-Q.

Costo por Usuario X_t^γ

Sea X_t^γ el costo de las atenciones de cada usuario que utilizará las tecnologías del conjunto Δ en el año τ con clasificación γ siendo $\gamma = 1$ para afiliados que nunca fueron infectados por covid-19 y $\gamma = 2$ para para afiliados que se recuperaron por covid-19.

Se estima el valor esperado de X_t^γ , asumiendo que el comportamiento de la variable está gobernado por un modelo probabilístico que captura la incertidumbre sobre su verdadero valor. Es decir que X_t^γ sigue una distribución de probabilidad continua con función de densidad $f(x; \theta)$, donde $x > 0$ y θ es un vector de parámetros desconocidos

Ajuste del modelo

Si el costo por usuario observado de los últimos doce meses, x_1, \dots, x_n , es una muestra de observaciones independientes de X_t , entonces se ajusta un modelo paramétrico conocido, $f(x; \theta)$, que maximiza la verosimilitud de la muestra de datos

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta), \quad (27)$$

en función de los parámetros del modelo. La expresión anterior se denomina función de verosimilitud y, aplicando logaritmo natural en ambos lados, se obtiene la función de log-verosimilitud:

$$l(\theta) = \sum_{i=1}^n \ln(f(x_i; \theta)). \quad (28)$$

Así pues, interesan los valores del vector de parámetros θ que maximizan el valor de la función $l(\theta)$, esto es:

$$\hat{\theta} = \operatorname{argmax}\{l(\theta)\}, \quad \theta \in \mathbb{R}^n \quad (29)$$

donde $\hat{\theta}$ es el vector de estimadores de máxima verosimilitud de θ . Finalmente, el modelo ajustado a los datos observados de los costos por usuario es la distribución de probabilidad con función de densidad $f(x; \hat{\theta})$ y vector de parámetros estimados $\hat{\theta}$.

Selección del modelo

En la ciencia estadística existen diferentes distribuciones de probabilidad discreta que se podrían ajustar a los datos observados de las utilidades por usuario. Suponga que existen n modelos rivales, con funciones de probabilidad, $f_1(x; \theta_1)$, $f_2(x; \theta_2)$, ..., $f_n(x; \theta_n)$, para $x > 0$ y θ_i es el vector de parámetros de la i -ésima función de probabilidad. La idea es replicar el procedimiento de maximización de la verosimilitud de los datos, estimación de

parámetros y ajuste de cada uno de los modelos rivales, con el fin de seleccionar el mejor entre ellos según el grado de ajuste a los datos.

Estadísticos de bondad de ajuste Basados en la verosimilitud.

Sea $\{X_n^r\}_{n < N}$ una muestra de tamaño N_r de la variable aleatoria que representa los valores utilidades por afiliado para la regional r , con función de densidad de probabilidad $f_{x,r}(x)$. La función de verosimilitud de la muestra observada para cada servicio regional, $L_r(\theta_r)$, está dada por la fórmula:

$$L_r(\theta_r) = \prod_{i=1}^{N_r} f_{x,r}(x_r; \theta_r), \tag{30}$$

donde θ_r es el vector de parámetros de la función de densidad de probabilidad para la regional r .

Ahora, si p es el número de parámetros de cada modelo para la regional r , los criterios de bondad de ajuste para la selección del mejor modelo son los siguientes:

- **Criterio de Akaike:** Ofrece una estimación relativa de la información perdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos.

$$AIC = -2 \ln(L_r(\theta_r)) + 2p. \tag{31}$$

- **Criterio de Akaike corregido:** El AICC penaliza los modelos más complejos que tienen un mayor número de parámetros, evitando así el sobreajuste. Se prefiere un modelo con un valor de AICC más bajo, lo que indica un mejor equilibrio entre el ajuste del modelo y la complejidad.

$$AICC = -2 \ln(L_r(\theta_r)) + \frac{2N_r p}{N_r - p - 1}. \tag{32}$$

- **Criterio de información bayesiano:** es una medida de bondad de ajuste de un modelo estadístico, y es utilizado como un criterio para para la selección de modelos entre un conjunto finito de modelos. Un BIC bajo implica un número menor de variables explicativas, mejor ajuste, o ambos.

$$BIC = -2 \ln(L_r(\theta_r)) + p \ln(N_r). \tag{33}$$

El mejor modelo es aquel con el menor valor del estadístico entre todos los criterios.

Estadísticos basados en la función de distribución empírica.

Esta clase de estadísticas se basa en la diferencia entre la estimación de la función de distribución acumulada (CDF) y la estimación de la función de distribución empírica (EDF). Sea $F_n(x)$ la función de distribución empírica y sea $F(x)$ la función de distribución acumulada.

- **Test de Kolmogorov-Smirnov:** se define como la distancia vertical máxima entre las funciones de distribución acumulada empíricas de dos muestras, o entre una función de distribución empírica y una función de distribución acumulada teórica de referencia. La ventaja principal de este estadístico es que es sensible a diferencias tanto en la localización como en la forma de la función de distribución acumulada:

$$KS = \max_{t \leq x \leq u} |F_n(x) - F(x)|, \quad (34)$$

donde t es un punto de truncamiento a izquierda y u es un punto de censura a la derecha.

- **Estadístico de Anderson Darling:** es un promedio ponderado de las diferencias al cuadrado entre la función empírica $F_n(x)$ y la función de distribución estimada $F(x)$:

$$AD = N \int_t^u \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]} f(x) dx. \quad (35)$$

Este estadístico tiende a poner más énfasis en el buen ajuste en las colas que en el medio de la distribución.

- **Cramer von Mises:** es un estadístico que es proporcional al valor esperado de la diferencia al cuadrado entre la función de distribución empírica y la función de distribución estimada.

$$CvM = N \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x). \quad (36)$$

Los criterios de selección consisten en ordenar y puntuar los n modelos rivales de acuerdo estadísticos bondad de ajuste basados en la verosimilitud, así como en aquellos basados en la distribución empírica. El mejor modelo es el que de acuerdo con los criterios anteriores tenga el mejor ajuste a los datos.

Una vez que se ranquean las distribuciones según el resultado de los criterios de información, con ayuda de los gráficos de contraste del Grafico Q-Q.

Capítulo III - Resultados

Análisis descriptivo – población objetivo.

En este apartado se profundizará en las características propias de la población afiliada a mutualser, la cual se ha recuperado del covid-19.

En primera instancia, se procede a observar los departamentos en los cuales se concentra la población objetivo. Para ello, se utiliza un diagrama de Pareto el cual permite observar de forma conjunta tanto el número de afiliados por departamento como su porcentaje de participación sobre el total.

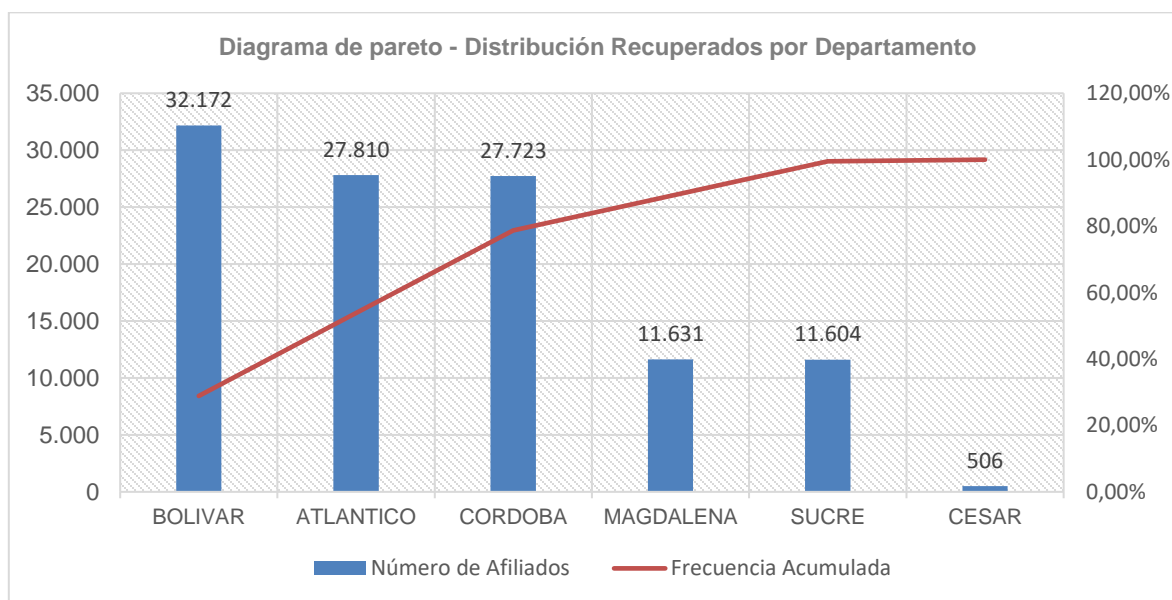
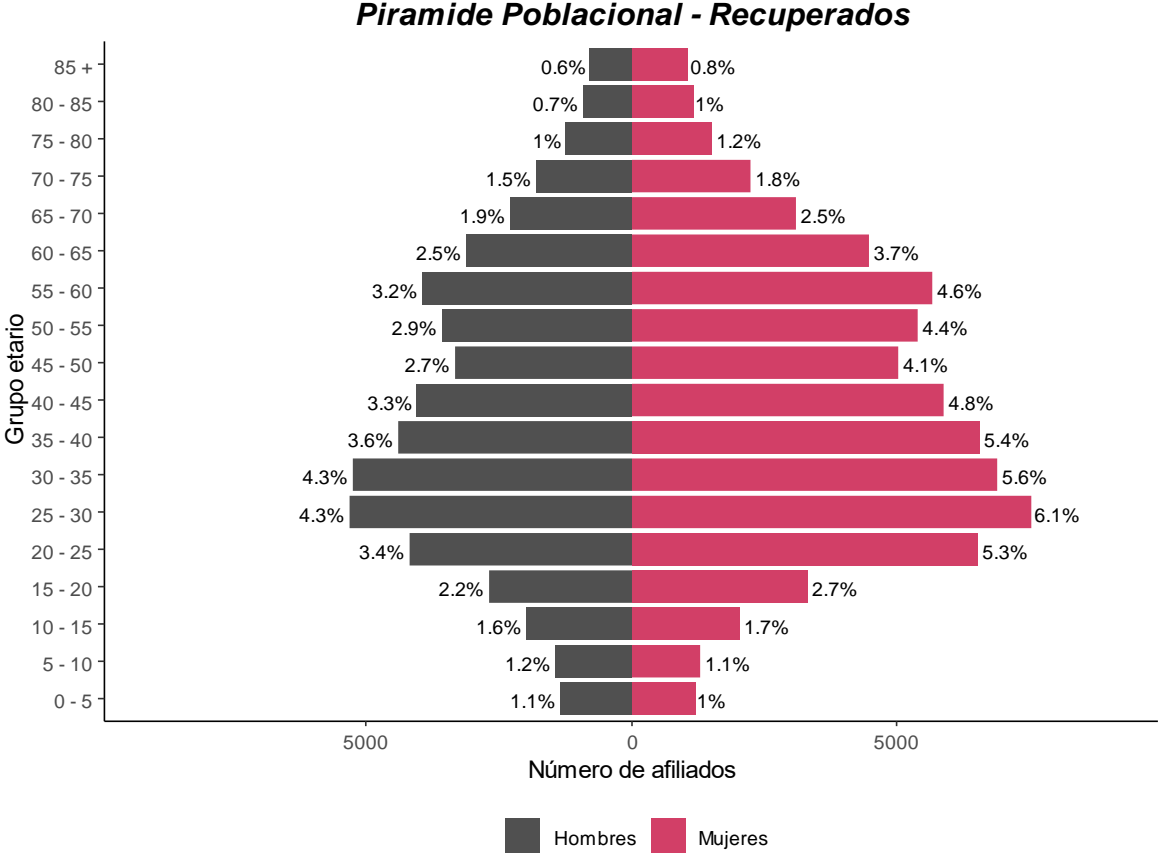


Figura 2. Diagrama de Pareto – Población recuperada del covid-19 por departamento
Fuente. Elaboración Propia.

Los resultados arrojaron que los departamentos Bolívar, Atlántico, Córdoba, Magdalena, Sucre y Cesar concentran más del 98% de la población recuperada, lo cual no es de extrañar, ya que, tal y como se mencionó anteriormente, es en estos donde opera la EPS de forma directa. Dicho esto, el análisis se restringirá a los 6 departamentos antes mencionados.

Previo al análisis de la población es importante dejar claro el tipo de población que compone el presente estudio, por ello, a continuación, se presenta la figura 3, la cual contiene pirámide poblacional de la población que se ha recuperado del covid-19.



*Figura 3. Pirámide poblacional – Población Recuperada del Covid-19
Fuente. Elaboración Propia.*

Analizando la figura 3 se logra observar una mayor concentración de mujeres en la población objetivo, así mismo se aprecia como del 55% de la población se encuentra concentrada en personas entre los 20 y 50 años, es decir, personas en edad productiva.

Adicionalmente, se introducen las figuras 4, y 5 las cuales contienen la pirámide poblacional para los departamentos de Atlántico, Bolívar, Córdoba; sucre, magdalena y cesar respectivamente.

Piramides Poblacionales - Recuperados de Covid-19

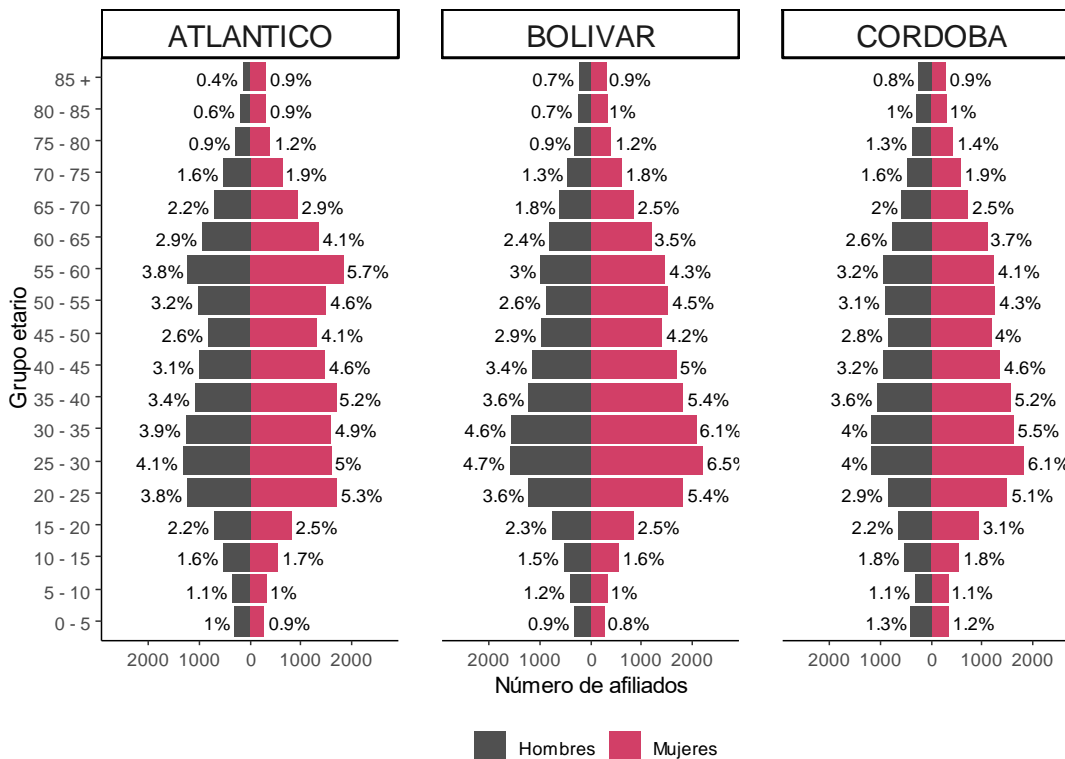


Figura 4. Pirámide poblacional de población recuperada del COVID 19 – Atlántico, Bolívar, y Córdoba
Fuente. Elaboración Propia.

Piramides Poblacionales - Recuperados de Covid-19

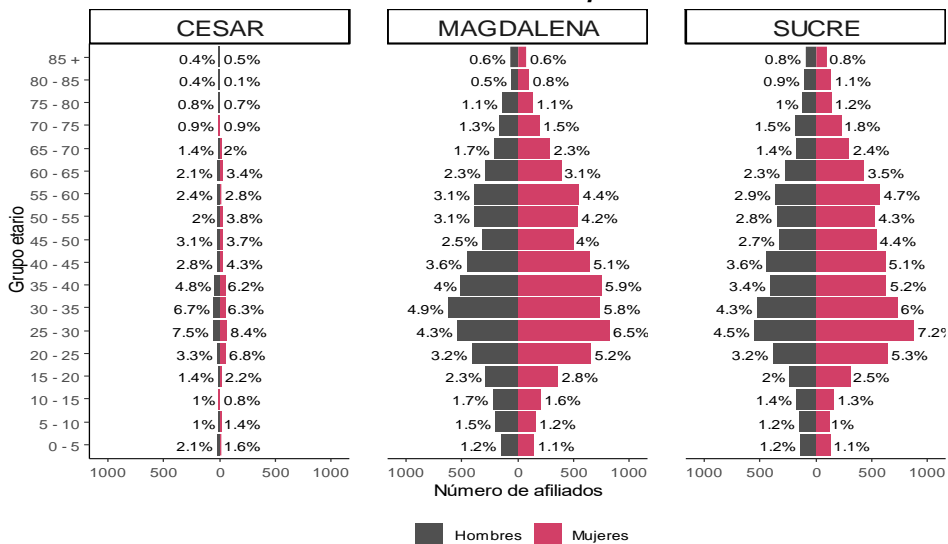


Figura 5. Pirámide poblacional de población recuperada del COVID 19 Cesar, Magdalena y Sucre.
Fuente. Elaboración Propia.

Tal y como se observa en las figuras 4 y 5, las poblaciones son similares dentro de las distintas regionales, siguiendo un comportamiento similar al de toda la población recuperada en donde las mujeres representan a la mayoría de la población objeto de estudio. Sin embargo, se observa como la regional atlántico posee un mayor porcentaje de personas recuperadas del covid-19 en el grupo de 55 a 60 años respecto al resto de regionales.

Síntomas PostCovid y nivel de calidad de vida.

Se procedió aplicar la encuesta construida con el fin de constatar cuales eran las condiciones de salud de la población objetivo frente a los síntomas PostCovid y su calidad de vida. Se trabajó con un listado de 111.147 afiliados, obteniendo los siguientes resultados.

Tipo Contacto	Cantidad	Participación
No efectivo	5395	73,18%
Efectivo	1977	26,82%
Encuesta realizada	1766	23,96%
Rechaza encuesta	211	2,86%
Total general	7372	100,00%

Tabla 1. Estadísticas Globales de encuesta Post COVID.
Fuente. Elaboración Propia.

Tal y como se logra apreciar, se contactó al 6,61% de la población objetivo, obteniendo una tasa de efectividad del 23,96%, el cual representa un contacto efectivo hacia el 1,58% de la población objetivo. La Tabla 2 presenta la distribución de los mismos a nivel regional.

Regional	Participación
Bolívar	26,84%
Sucre	15,40%
Magdalena	15,18%
Atlántico	14,67%
Cesar	14,16%
Córdoba	13,65%
Total general	100,00%

Tabla 2. Distribución Regional de encuestas Post COVID realizadas.
Fuente. Elaboración Propia.

Al analizar los resultados de la encuesta se obtiene que el porcentaje de personas a que manifestaron al menos un síntoma Postcovid fue del 68,21%. Por otra parte, el 51,08% de la población encuestada manifestó tener al menos una dificultad en su calidad de vida. Las

tablas 3 y 4 presentan la distribución de las respuestas de los bloques “Post Covid” y “Calidad de Vida” respectivamente para aquellos encuestados que manifestaron tener al menos una condición o síntoma.

Pregunta	Respuesta	Proporción
¿una vez recuperado del covid-19 ha presentado dificultades para dormir?	OCASIONALMENTE	56,44%
¿una vez recuperado del covid-19 ha presentado dificultades para dormir?	CASI SIEMPRE	35,21%
¿una vez recuperado del covid-19 ha presentado dificultades para dormir?	CASI NUNCA	5,63%
¿una vez recuperado del covid-19 ha presentado dificultades para dormir?	SIEMPRE	2,72%
¿una vez recuperado del covid-19 ha presentado pérdida parcial o total del sentido del olfato?	OCASIONALMENTE	69,12%
¿una vez recuperado del covid-19 ha presentado pérdida parcial o total del sentido del olfato?	CASI NUNCA	15,69%
¿una vez recuperado del covid-19 ha presentado pérdida parcial o total del sentido del olfato?	CASI SIEMPRE	13,73%
¿una vez recuperado del covid-19 ha presentado pérdida parcial o total del sentido del olfato?	SIEMPRE	1,47%
¿una vez recuperado del covid-19 ha sufrido usted de dificultad respiratoria?	OCASIONALMENTE	64,66%
¿una vez recuperado del covid-19 ha sufrido usted de dificultad respiratoria?	CASI SIEMPRE	26,15%
¿una vez recuperado del covid-19 ha sufrido usted de dificultad respiratoria?	CASI NUNCA	8,62%
¿una vez recuperado del covid-19 ha sufrido usted de dificultad respiratoria?	SIEMPRE	0,57%
¿una vez recuperado del covid-19 ha sufrido usted de dolor de cabeza frecuente?	OCASIONALMENTE	55,36%
¿una vez recuperado del covid-19 ha sufrido usted de dolor de cabeza frecuente?	CASI SIEMPRE	36,16%
¿una vez recuperado del covid-19 ha sufrido usted de dolor de cabeza frecuente?	CASI NUNCA	6,25%
¿una vez recuperado del covid-19 ha sufrido usted de dolor de cabeza frecuente?	SIEMPRE	2,23%
¿una vez recuperado del covid-19 ha sufrido usted de dolor en el pecho frecuente?	OCASIONALMENTE	66,49%
¿una vez recuperado del covid-19 ha sufrido usted de dolor en el pecho frecuente?	CASI SIEMPRE	21,08%
¿una vez recuperado del covid-19 ha sufrido usted de dolor en el pecho frecuente?	CASI NUNCA	11,08%
¿una vez recuperado del covid-19 ha sufrido usted de dolor en el pecho frecuente?	SIEMPRE	1,35%
¿una vez recuperado del covid-19 ha sufrido usted de fatiga?	OCASIONALMENTE	68,14%

¿una vez recuperado del covid-19 ha sufrido usted de fatiga?	CASI SIEMPRE	26,58%
¿una vez recuperado del covid-19 ha sufrido usted de fatiga?	CASI NUNCA	3,16%
¿una vez recuperado del covid-19 ha sufrido usted de fatiga?	SIEMPRE	2,11%
¿una vez recuperado del covid-19 ha sufrido usted de problemas estomacales frecuentes?	OCASIONALMENTE	61,71%
¿una vez recuperado del covid-19 ha sufrido usted de problemas estomacales frecuentes?	CASI SIEMPRE	28,29%
¿una vez recuperado del covid-19 ha sufrido usted de problemas estomacales frecuentes?	CASI NUNCA	7,80%
¿una vez recuperado del covid-19 ha sufrido usted de problemas estomacales frecuentes?	SIEMPRE	2,20%

*Tabla 3. Distribución de respuestas afirmativas – Bloque de preguntas Post COVID.
Fuente. Elaboración Propia.*

Pregunta	Respuesta	Proporción
¿fue usted capaz de retomar sus actividades cotidianas (trabajo y demás actividades)?	Si, pero con ligeras molestias	72,39%
¿fue usted capaz de retomar sus actividades cotidianas (trabajo y demás actividades)?	Si, pero con muchas molestias	20,25%
¿fue usted capaz de retomar sus actividades cotidianas (trabajo y demás actividades)?	No, no me fue posible	7,36%
¿una vez recuperado del covid-19 ha sufrido ansiedad?	Si, Me he sentido ligeramente ansioso(a)	55,53%
¿una vez recuperado del covid-19 ha sufrido ansiedad?	Si, Me he sentido bastante ansioso(a)	44,47%
¿una vez recuperado del covid-19 ha sufrido depresión?	Si, Me he sentido ligeramente deprimido(a)	58,92%
¿una vez recuperado del covid-19 ha sufrido depresión?	Si, Me he sentido bastante deprimido(a)	41,08%
¿una vez recuperado del covid-19, es usted capaz de caminar?	Si, pero con ligeras molestias	76,33%
¿una vez recuperado del covid-19, es usted capaz de caminar?	Si, pero con muchas molestias	17,39%
¿una vez recuperado del covid-19, es usted capaz de caminar?	No, no me es posible.	6,28%
¿una vez recuperado del covid-19, ha sufrido de algún tipo de dolencias?	Si, he tenido dolores leves	36,16%
¿una vez recuperado del covid-19, ha sufrido de algún tipo de dolencias?	Si, sufro de dolor constante	32,18%
¿una vez recuperado del covid-19, ha sufrido de algún tipo de dolencias?	Si, he sufrido dolores moderados	31,66%

Tabla 4. Distribución de respuestas afirmativas – Bloque de preguntas Calidad de Vida.
Fuente. Elaboración Propia.

De acuerdo con los datos presentados en las tablas 3 y 4, se observa que la mayoría de las personas que experimentan al menos un síntoma PostCovid y/o que afecta su calidad de vida lo hacen de manera leve u ocasional (Sombreadas en Verde). No obstante, en la mayoría de los casos, las respuestas con la segunda opción de mayor participación en cada pregunta corresponden a frecuencias altas e intensidades mayores. (Sombreadas en Amarillo)

Adicionalmente se procede a comparar la presencia de síntomas en las distintas regionales para cada régimen, las figuras 3 y 4 presentan la comparativa en los bloques “Post Covid” y “Calidad de Vida” respectivamente

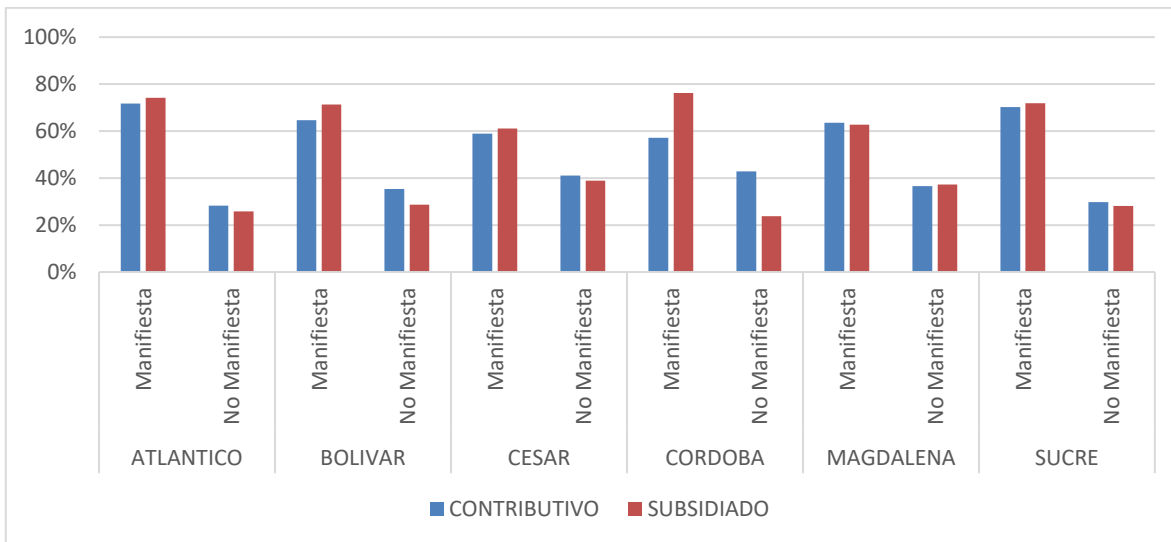


Figura 6. Comparativo - Presencia de síntomas Post COVID por régimen y regional.
Fuente. Elaboración Propia.

La figura 6 sugiere que en la mayoría de los casos no se presenta una diferencia notoria a nivel de cada régimen dentro de las distintas regionales ya que las proporciones están muy parejas. Sin embargo, se observa el caso particular de la regional de Córdoba, en donde el 76% de los afiliados del régimen subsidiado manifiestan al menos un síntoma Post COVID, mientras que en el régimen contributivo dicha proporción se limita al 57%

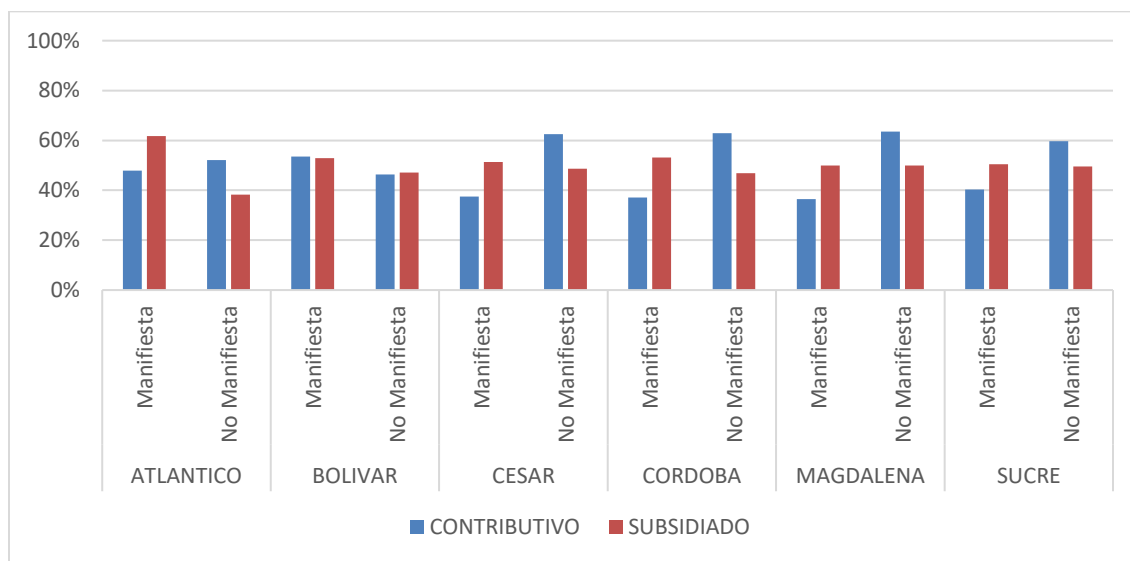


Figura 7. Comparativo - Dificultades en la calidad de vida por régimen y regional.
Fuente. Elaboración Propia.

Por su parte, la figura 7 sugiere que, con excepción de la regional Bolívar, todas las regionales presentan diferencias notorias en la ocurrencia de síntomas que afectan su

calidad de vida, siendo los afiliados del régimen subsidiado aquellos que mayormente manifiestan los síntomas antes mencionados.

Adicionalmente es importante tener presente el tipo de población que se encuentra manifestando dichas afectaciones con el fin de tener un panorama mucho mas claro de las posibles afectaciones y riesgos en salud asociados. Las Figuras 8 y 9 Presentan las pirámides poblacionales de la población que ha manifestado afectaciones a su calidad de vida y síntomas PostCovid Respectivamente.

Incremento en los costos de atención (PreCovid Vs PostCovid)

En esta sección, se presentan los incrementos del costo medio de atención en dos períodos específicos: "Pre Covid" y "Post Covid". Estos períodos fueron definidos previamente acorde a la ecuación (17). La intención de este análisis es comparar y contrastar los costos promedio asociados a la atención de los afiliados que se han recuperado del COVID en ambos períodos utilizando los años 2019 y 2022 cuyos valores se han indexado a 2023 mediante inflación.

A continuación, se presenta la tabla 5, donde se detalla el incremento porcentual que ha tenido el costo mensual por persona en 2022 frente a 2019.

Mes	Incremento
ENERO	22,32%
FEBRERO	29,27%
MARZO	72,47%
ABRIL	75,66%
MAYO	75,33%
JUNIO	21,71%
JULIO	65,04%
AGOSTO	51,21%

Tabla 5. Incremento de Costo medio de Atención PreCovid Vs PostCovid (2019 Vs 2022)
Fuente. Elaboración Propia.

Se observa como los meses desde marzo hasta mayo presentan los mayores incrementos, los cuales fueron del 72,47%, 75,66% y 75,33%, respectivamente. Sin embargo, a partir del mes de junio los datos pueden sugerir un cambio de tendencia.

Lo siguiente es analizar el comportamiento de dichos incrementos al interior de cada regional. La figura 5 ilustra el incremento de porcentual de la atención PreCovid Vs PostCovid en cada una de las regionales.

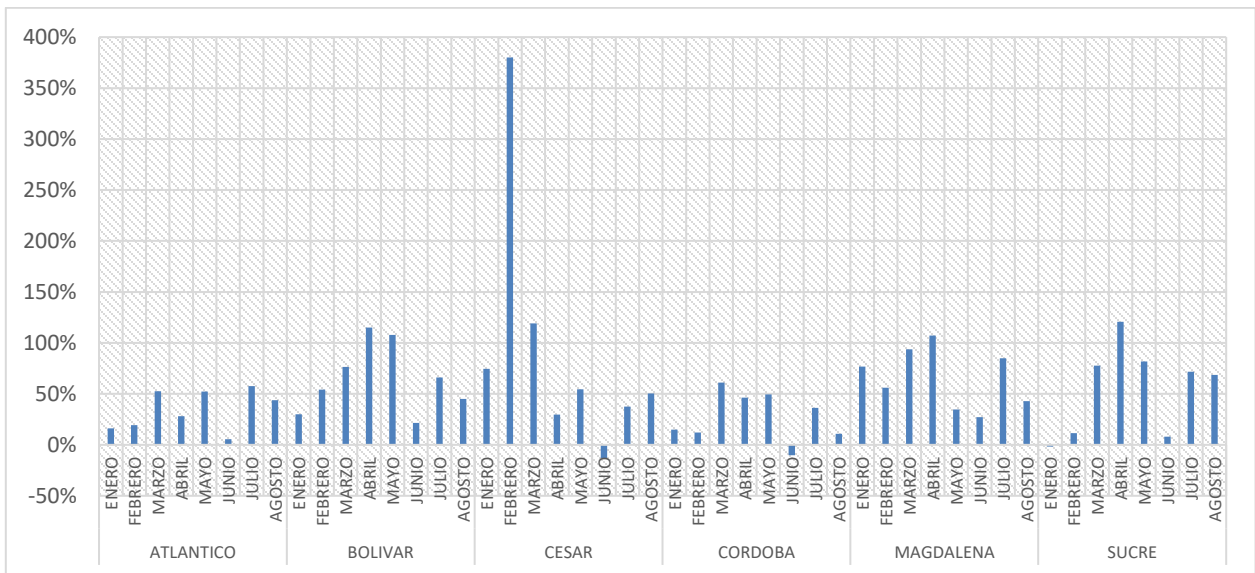


Figura 8. Incremento de Costo medio de Atención PreCovid Vs PostCovid (2019 Vs 2022) Por Regional.
Fuente. Elaboración Propia.

Tal y como se logra apreciar las regionales Cesar, Bolívar, Magdalena y Sucre son aquellas que presentan los incrementos más altos con un promedio de 92%, 65%, 65% y 55% respectivamente. Así mismo, se logra que el mes de abril es donde se presentan los mayores incrementos en dichas regionales.

Con base al análisis realizado, se puede concluir que efectivamente existe un aumento considerable del costo de servicios en salud asociado a aquellas personas que superaron el covid-19. Dicho aumento corresponde en promedio a un 51,63%.

Así mismo, el hecho de que este fenómeno se haya observado en los distintos departamentos de forma consistente demuestra que el impacto económico del covid-19 no está limitado a la fase aguda de la infección, sino que se extiende posterior a la recuperación.

Incremento en el número de utilizaciones (PreCovid Vs PostCovid)

En esta sección, se presentan los incrementos las utilizaciones promedio por usuario en dos períodos específicos: "Pre Covid" y "Post Covid". Estos períodos fueron definidos previamente acorde a la ecuación 11. La intención de este análisis es comparar y contrastar el número de utilizaciones promedio que demandan los afiliados que se han recuperado del COVID en ambos períodos utilizando los años 2019 y 2022.

Mes	Incremento
ENERO	-21,56%
FEBRERO	3,57%
MARZO	1,75%
ABRIL	10,65%
MAYO	1,35%
JUNIO	24,72%
JULIO	-10,24%
AGOSTO	0,25%

Tabla 6. Incremento del número de utilizaciones por usuario – Pre COVID Vs Post COVID (2019 Vs 2022)
Fuente. Elaboración Propia.

La tabla 7 sugiere un incremento considerable en las utilizaciones a través de los distintos meses, siendo abril y junio aquellos con un mayor incremento.

A continuación, se presenta la figura 10, la cual consolida los incrementos en el número de utilizaciones promedio afiliado para cada regional.

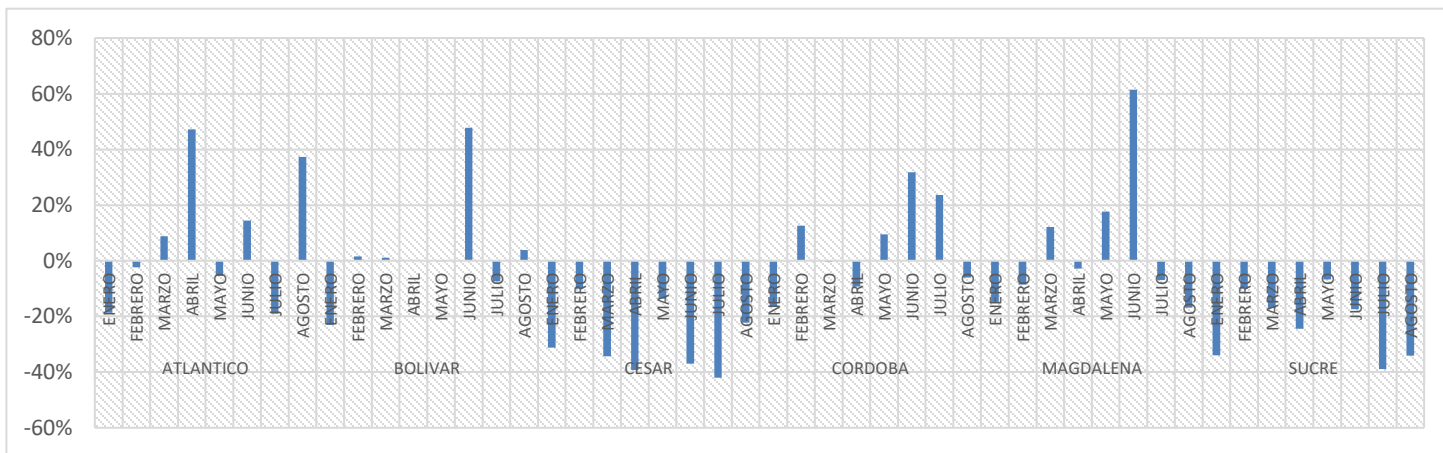


Figura 9. Incremento de las utilizaciones promedio por usuario PreCovid Vs PostCovid (2019 Vs 2022) Por Regional.
Fuente. Elaboración Propia.

Al analizar el incremento de las utilizaciones a nivel departamental, se aprecia como estas han disminuido en gran mayoría, no obstante, se aprecia que existen meses tales como junio en los cuales se ha presentado un incremento notable. Dicho esto, el incremento en

el número de utilizations promedio por usuario Recuperado del covid-19 en cada regional queda de la siguiente manera.

Regional	Incremento Promedio
Atlántico	7,77%
Bolívar	2,99%
Cesar	-28,71%
Córdoba	5,68%
Magdalena	5,15%
Sucre	-22,84%

*Tabla 7. Incremento Promedio en las utilizations (Pre Covid Vs Post Covid) (2019 Vs 2022)
Fuente. Elaboración Propia.*

Dicho lo anterior se puede concluir que en promedio las personas que se han recuperado del covid 19 tienen mayor frecuencia de uso de servicios de salud en 2022 respecto a 2019

Utilizaciones esperadas 2023

A continuación, se procede a estimar el número de utilizaciones esperadas en el año 2023 para los usuarios recuperados del covid-19 y aquellos que no fueron contagiados por dicha enfermedad mediante el ajuste de funciones de densidad de probabilidad.

Ajuste de funciones de densidad de probabilidad – Afiliados no infectados por covid-19

Regional Atlántico.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Atlántico es el siguiente:

Distribución de probabilidad	Kolmogorov-Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.0648	703.6663	4336429	4336429	4336440
Poisson	0.5273	39656.1842	21866134	21866134	21866145
Binomial Negativa Tipo I	0.0829	273.3673 *	4326293	4326293	4326315
Logarítmica	0.2536	12327.9687	4413639	4413639	4413650
Poisson Generalizada	0.0518 *	358.071	4297985 *	4297985 *	4298007 *

Tabla 8. Criterios de Ajuste – $u_{i,t}^{(4)}$ Regional Atlántico.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Poisson Generalizada y la Binomial negativa tipo I.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

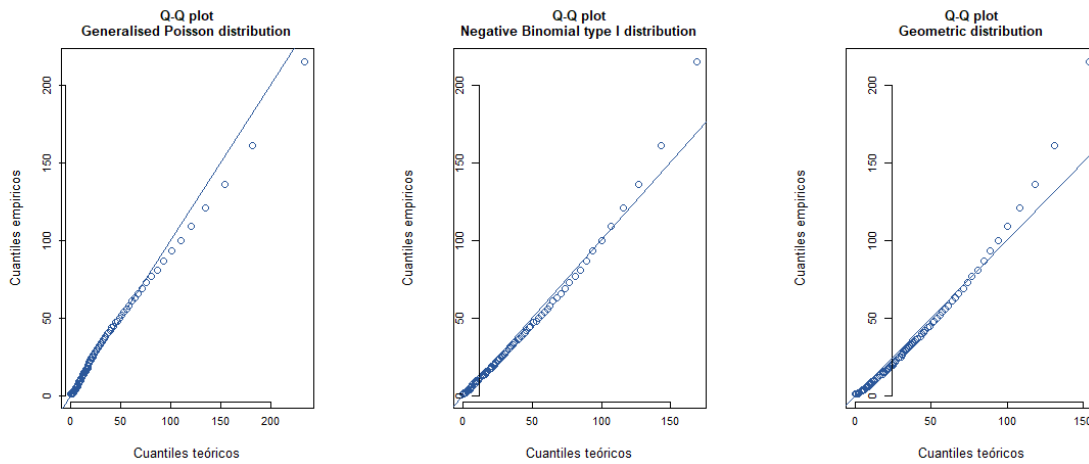


Figura 10. Gráficos Q-Q – $u_{i,t}^{(1)}$ Regional Atlántico.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	33.1240333	32.9898730	33.2587391
sigma	0.2191285	0.2185259	0.2197327

Tabla 9. Parámetros del modelo seleccionado - Regional Atlántico.
Fuente. Elaboración Propia.

Regional Bolívar.

El resumen sobre el ajuste de las distribuciones – consideradas para la Regional Bolívar es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.1088	2764.8666	5583618	5583618	5583629
Poisson	0.5322	59083.0791	24328963	24328963	24328974
Binomial Negativa Tipo I	0.1363	1232.2697	5549091	5549091	5549114
Logarítmica	0.2157	9645.4885	5440724 *	5440724 *	5440735 *
Poisson Generalizada	0.1065 *	411.882 *	5453651	5453651	5453674

Tabla 10. Criterios de Ajuste – $u_{14}^{(1)}$ Regional Bolívar.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Poisson Generalizada y la Logarítmica.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

Los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos son:

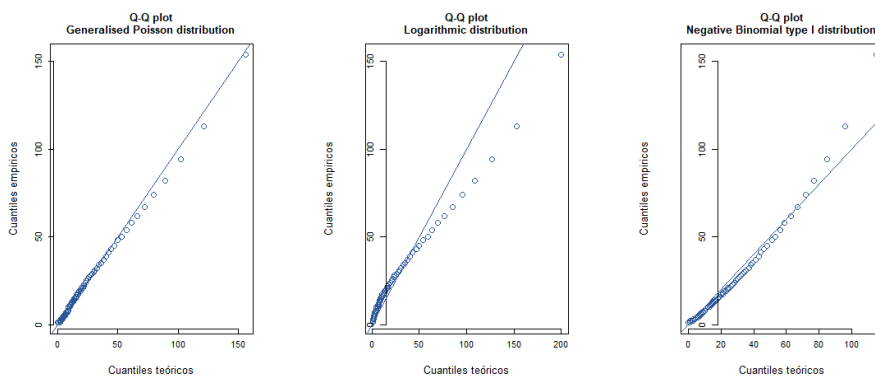


Figura 11. Gráficos Q-Q – $u_{14}^{(1)}$ Regional Bolívar.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	21.2413379	21.1660794	21.3168641

Parámetro	Estimación	2.5 %	97.5 %
sigma	0.2779829	0.2773151	0.2786524

Tabla 11. Parámetros del modelo seleccionado - Regional Bolívar.
Fuente. Elaboración Propia.

Regional Córdoba.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Córdoba es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.0879	1270.5037	4254645	4254645	4254656
Poisson	0.5233	40891.3322	18951862	18951862	18951873
Binomial Negativa Tipo I	0.1108	448.5939	4236309	4236309	4236332
Logarítmica	0.2098	9034.757	4231002	4231002	4231013
Poisson Generalizada	0.079 *	353.1564 *	4197530 *	4197530 *	4197552 *

Tabla 12. Criterios de Ajuste - $u_{i,t}^{(1)}$ Regional Córdoba.
Fuente. Elaboración Propia.

Según los criterios de información, la distribución que presenta mejor ajuste es la Poisson Generalizada.

Sin embargo, previo a seleccionar esta como la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos para confirmar la decisión.

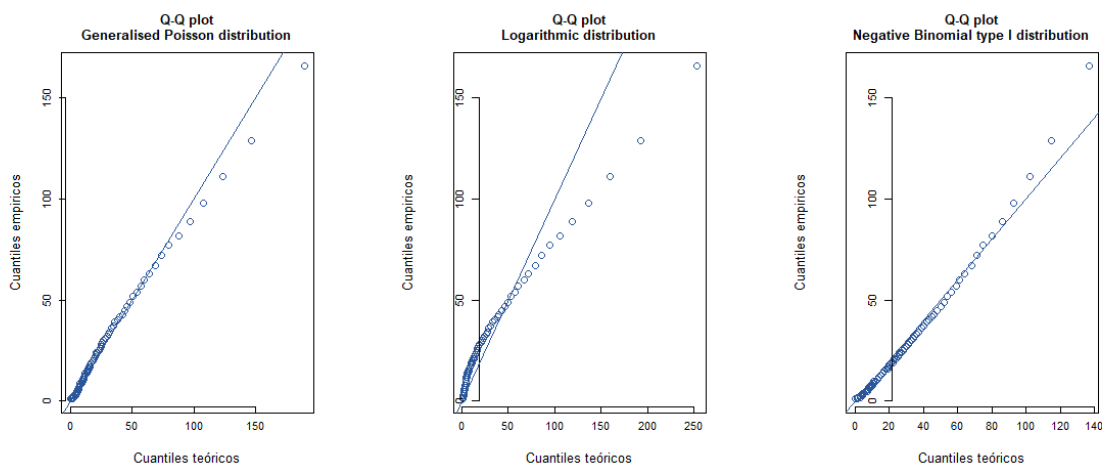


Figura 12. Gráficos Q-Q - $u_{i,t}^{(1)}$ Regional Córdoba.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	25.9420583	25.835284	26.0492737
sigma	0.2529007	0.252199	0.2536043

Tabla 13. Parámetros del modelo seleccionado - Regional Córdoba.
Fuente. Elaboración Propia.

Regional Magdalena.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Magdalena es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.0915	351.3386	1655244	1655244	1655254
Poisson	0.4946	15384.3073	6302747	6302747	6302757
Binomial Negativa Tipo I	0.1189	230.2602	1651877	1651877	1651898
Logarítmica	0.2235	3860.6846	1646141	1646141	1646151
Poisson Generalizada	0.0901 *	120.199 *	1630052 *	1630052 *	1630072 *

Tabla 14. Criterios de Ajuste – $u_{it}^{(1)}$ Regional Magdalena.
Fuente. Elaboración Propia.

Según los criterios de información, la distribución que presenta mejor ajuste es la Poisson Generalizada.

Sin embargo, previo a seleccionar esta como la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos para confirmar la decisión.

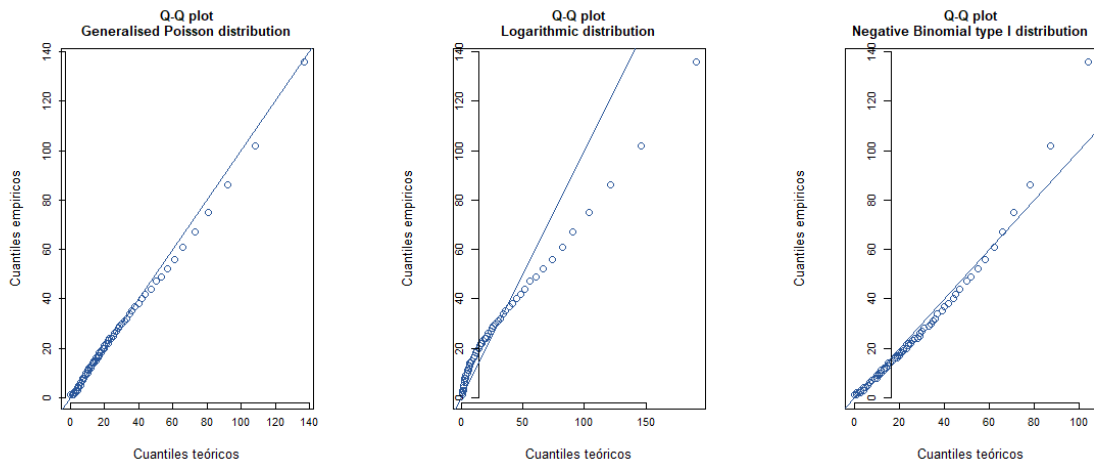


Figura 13. Gráficas Q-Q – $u_{i,t}^{(1)}$ Regional Magdalena.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	20.3474920	20.2271402	20.468560
sigma	0.2546655	0.2535371	0.255799

Tabla 15. Parámetros del modelo seleccionado - Regional Magdalena.
Fuente. Elaboración Propia.

Regional Sucre.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Sucre es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.0801	406.7433	1920787	1920787	1920797
Poisson	0.507	17803.2738	7939157	7939157	7939167
Binomial Negativa Tipo I	0.1104	203.8717	1915689	1915689	1915710
Logarítmica	0.22	4495.8366	1917946	1917946	1917956
Poisson Generalizada	0.0797 *	147.3586 *	1895290 *	1895290 *	1895311 *

Tabla 16. Criterios de Ajuste – $u_{i,t}^{(1)}$ Regional Sucre.
Fuente. Elaboración Propia.

Según los criterios de información, la distribución que presenta mejor ajuste es la Poisson Generalizada.

Sin embargo, previo a seleccionar esta como la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos para confirmar la decisión.

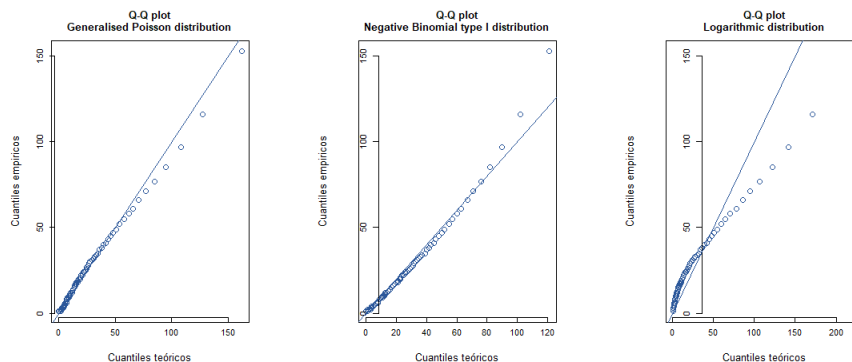


Figura 14. Gráficos Q-Q – $\mu_{14}^{(1)}$ Regional Sucre.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	23.4435016	23.3086492	23.5791343
sigma	0.2488795	0.2478539	0.2499094

Tabla 17. Parámetros del modelo seleccionado - Regional Sucre.
Fuente. Elaboración Propia.

Regional Cesar.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Cesar es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.1485	76.7536	91089	91089	91096
Poisson	0.533	1041.0825	368761	368761	368768
Binomial Negativa Tipo I	0.1743	39.133	90232	90232	90247
Logarítmica	0.2308	109.7058	85927 *	85927 *	85935 *

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Poisson Generalizada	0.1524	16.8973 *	88040	88040	88055

Tabla 18. Criterios de Ajuste – $u_{i,t}^{(1)}$ Regional Cesar.

Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Poisson Generalizada y la Logarítmica.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

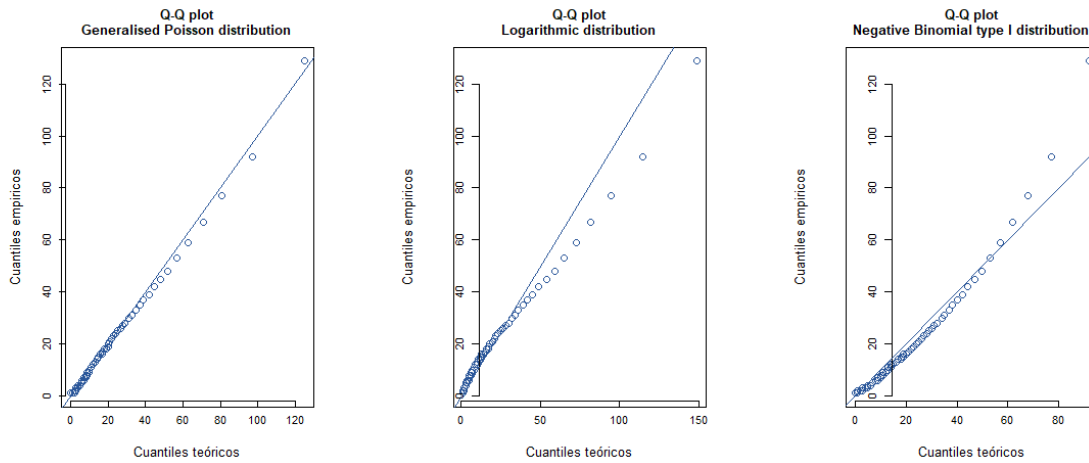


Figura 15. Gráficos Q-Q – $u_{i,t}^{(1)}$ Regional Cesar.

Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	16.5426409	16.0909343	17.0070279
sigma	0.3179871	0.3120884	0.3239972

Tabla 19. Parámetros del modelo seleccionado - Regional Bolívar.

Fuente. Elaboración Propia.

Ajuste de funciones de densidad de probabilidad – Afiliados Recuperados del Covid-19

Regional Atlántico.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Atlántico es el siguiente:

Distribución de probabilidad	Kolmogorov-Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.0457 *	23.4458	288041	288041	288049
Poisson	0.5335	2521.1185	1708536	1708536	1708544
Binomial Negativa Tipo I	0.0563	13.2297 *	287791	287791	287808
Logarítmica	0.2991	1055.5666	299835	299835	299843
Poisson Generalizada	0.0591	28.4255	286617 *	286617 *	286633 *

Tabla 20. Criterios de Ajuste – $u_{i,t}^{(2)}$ Regional Atlántico.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Poisson Generalizada y la Geométrica.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

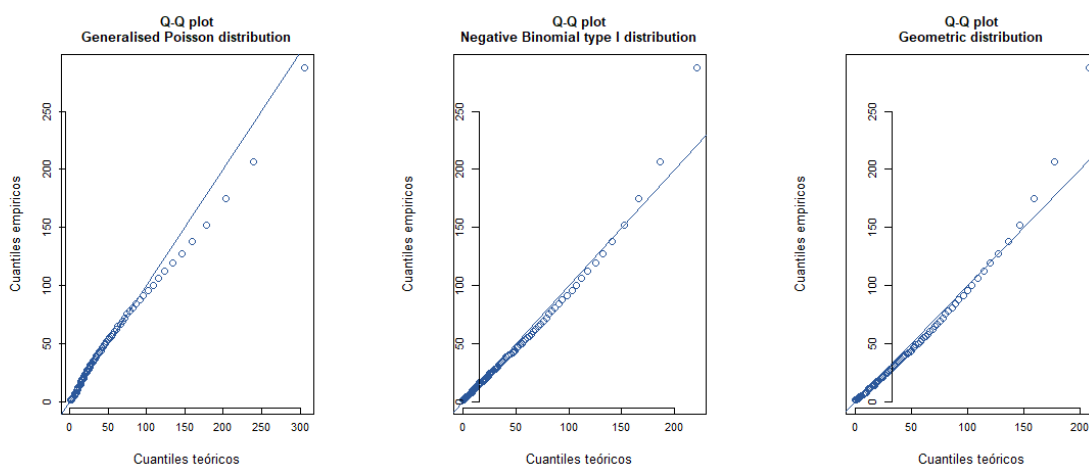


Figura 16. Gráficos Q-Q – $u_{i,t}^{(2)}$ Regional Atlántico.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	44.7874399	44.090336	45.4955657
sigma	0.1845521	0.182591	0.1865343

Tabla 21. Parámetros del modelo seleccionado - Regional Bolívar.
Fuente. Elaboración Propia.

Regional Bolívar.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Bolívar es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.0638	50.766	287764	287764	287773
Poisson	0.5332	2694.5212	1596974	1596974	1596982
Binomial Negativa Tipo I	0.0756	27.9034	287251	287251	287268
Logarítmica	0.274	910.2882	294957	294957	294965
Poisson Generalizada	0.044 *	14.9396 *	284141 *	284141 *	284158 *

Tabla 22. Criterios de Ajuste – $u_{i,2}^{(2)}$ Regional Bolívar.
Fuente. Elaboración Propia.

Según los criterios de información, la distribución que presenta mejor ajuste es la Poisson Generalizada.

Sin embargo, previo a seleccionar esta como la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos para confirmar la decisión.

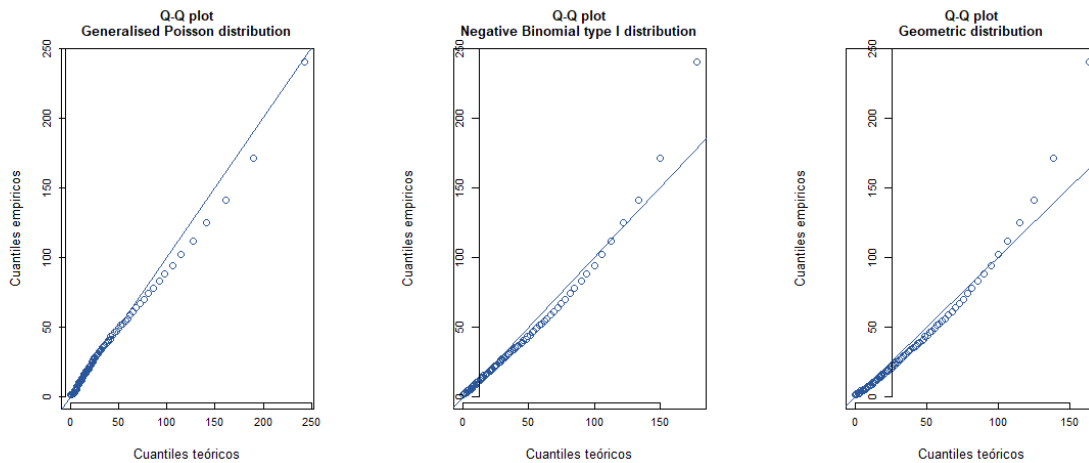


Figura 17. Gráficos Q-Q – $u_{t,t}^{(2)}$ Regional Bolívar.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	35.2239757	34.6833632	35.7730149
sigma	0.2073889	0.2052014	0.2095996

Tabla 23. Parámetros del modelo seleccionado - Regional Bolívar.
Fuente. Elaboración Propia.

Regional Córdoba.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Córdoba es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Anderson-Darling	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.056 *	184.765	30.2052	259661	259661	259669
Poisson	0.536	Inf	2270.5293	1513205	1513205	1513213
Binomial Negativa Tipo I	0.0643	110.2225 *	10.2064 *	259207	259207	259224
Logarítmica	0.2775	4059.852	849.1935	268019	268019	268027
Poisson Generalizada	0.0602	149.0509	28.9133	258330 *	258330 *	258346 *

Tabla 24. Criterios de Ajuste – $u_{t,t}^{(2)}$ Regional Córdoba.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Poisson Generalizada y la Binomial Negativa Tipo I.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

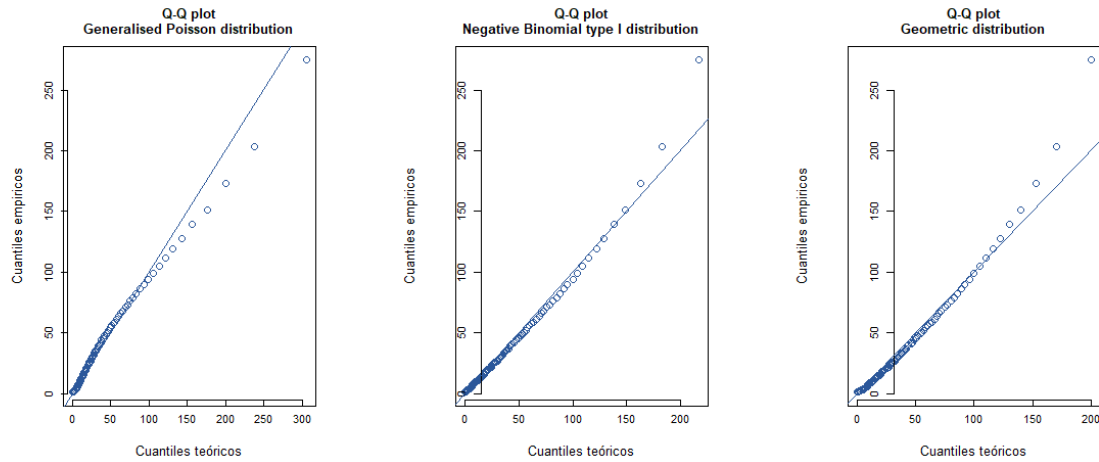


Figura 18. Gráficos Q-Q – $u_{12}^{(2)}$ Regional Córdoba.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	43.063322	42.3318725	43.8074099
sigma	0.196426	0.1942244	0.1986525

Tabla 25. Parámetros del modelo seleccionado - Regional Bolívar.
Fuente. Elaboración Propia.

Regional Magdalena.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Magdalena es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.0656	17.9803	103238	103238	103245
Poisson	0.5304	984.549	533866	533866	533873
Binomial Negativa Tipo I	0.0801	12.1926	103100	103100	103115
Logarítmica	0.2751	330.7527	105659	105659	105667
Poisson Generalizada	0.0446 *	4.3652 *	101764 *	101764 *	101779 *

Tabla 26. Criterios de Ajuste – $u_{i,t}^{(2)}$ Regional Magdalena.
Fuente. Elaboración Propia.

Según los criterios de información, la distribución que presenta mejor ajuste es la Poisson Generalizada.

Sin embargo, previo a seleccionar esta como la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos para confirmar la decisión.

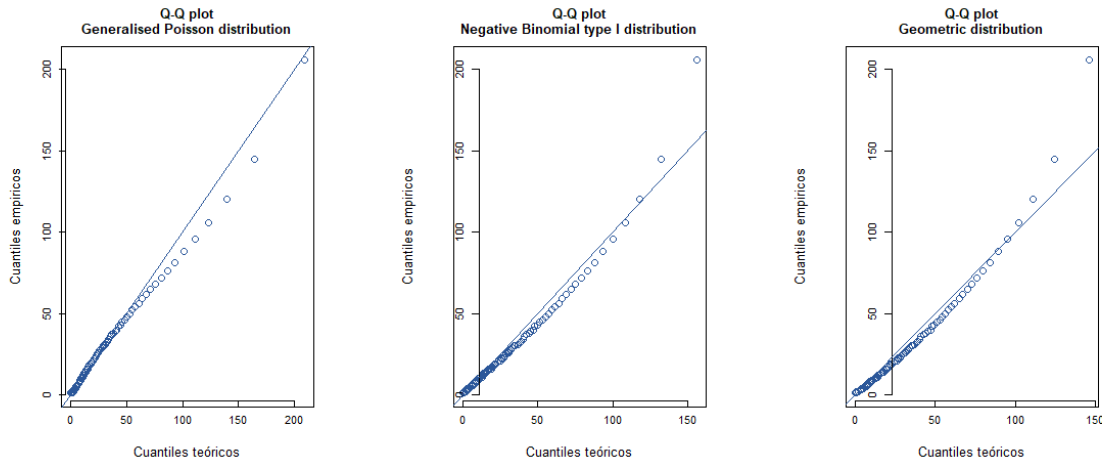


Figura 19. Gráficos Q-Q – $u_{i,t}^{(2)}$ Regional Magdalena.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	31.2758995	30.5105904	32.0604052
sigma	0.2112137	0.2075146	0.2149788

Tabla 27. Parámetros del modelo seleccionado - Regional Magdalena.
Fuente. Elaboración Propia.

Regional Sucre.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Sucre es el siguiente:

Distribución de probabilidad	Kolmogorov-Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.0627	17.297	102275	102275	102282
Poisson	0.5344	946.7619	582213	582213	582220
Binomial Negativa Tipo I	0.0776	8.7932	102041	102041	102056
Logarítmica	0.2678	314.1386	104515	104515	104522
Poisson Generalizada	0.0554 *	8.4675 *	101072 *	101072 *	101086 *

Tabla 28. Criterios de Ajuste – $u_{i,t}^{(2)}$ Regional Sucre.
Fuente. Elaboración Propia.

Según los criterios de información, la distribución que presenta mejor ajuste es la Poisson Generalizada.

Sin embargo, previo a seleccionar esta como la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos para confirmar la decisión.

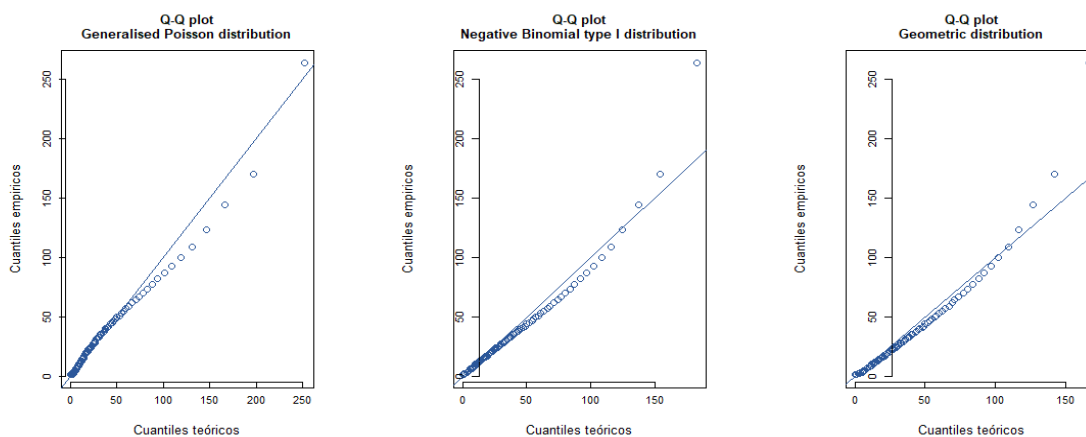


Figura 20. Gráficos Q-Q – $u_{i,t}^{(2)}$ Regional Sucre.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	35.9127506	34.9737655	36.8769460
sigma	0.2101161	0.2063949	0.2139044

Tabla 29. Parámetros del modelo seleccionado - Regional Sucre.
Fuente. Elaboración Propia.

Regional Cesar.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Cesar es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Geométrica	0.1488	3.4355	4772	4772	4777
Poisson	0.5594	54.0449	22398	22398	22402
Binomial Negativa Tipo I	0.1384	1.6157	4740	4740	4748
Logarítmica	0.216	8.1722	4639	4639	4644
Poisson Generalizada	0.1064 *	0.3704 *	4634 *	4634 *	4642 *

Según los criterios de información, la distribución que presenta mejor ajuste es la Poisson Generalizada.

Sin embargo, previo a seleccionar esta como la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos para confirmar la decisión.

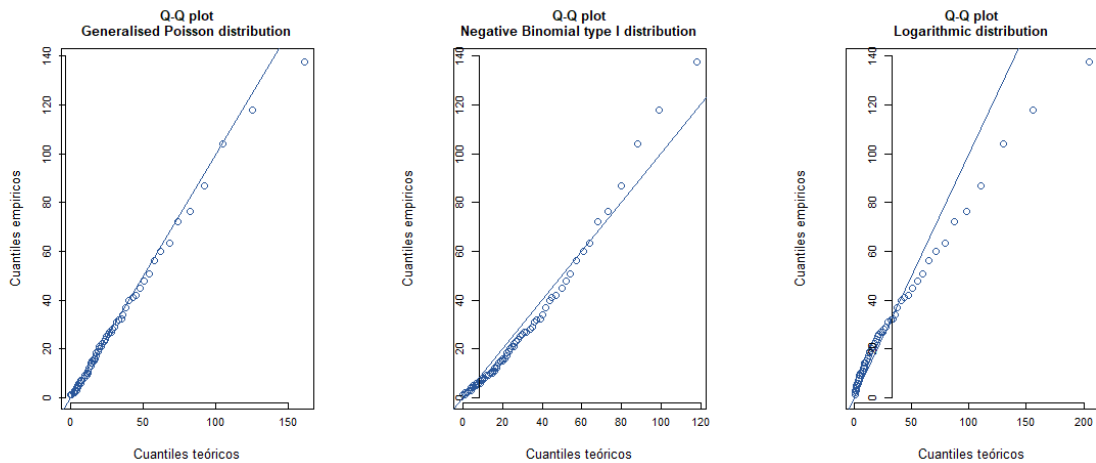


Figura 21. Criterios de Ajuste – $u_{14}^{(2)}$ Regional Cesar.
Fuente. Elaboración Propia.

Se observa como la distribución Poisson Generalizada Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	21.6580757	19.1526152	24.4912894
sigma	0.2789854	0.2571476	0.3026778

Tabla 30. Parámetros del modelo seleccionado - Regional Sucre.
Fuente. Elaboración Propia.

Resumen - Utilizaciones Esperadas 2023 (Usuarios Recuperados del Covid Vs No Infectados)

Una vez se han ajustado las distribuciones de probabilidad, se procede a calcular el incremento en las utilizaciones esperadas para el año 2023. La Tabla 34 presenta los incrementos de las utilizaciones esperadas de los usuarios recuperados por covid frente a aquellos que nunca fueron infectados por dicho virus.

Regional	Distribución Seleccionada (Afiliados Recuperados del COVID)	Distribución Seleccionada (Afiliados no Infectados por COVID)	Incremento en las utilizaciones
Atlántico	Poisson Generalizada	Poisson Generalizada	35,25%
Bolívar	Poisson Generalizada	Poisson Generalizada	66,77%
Córdoba	Poisson Generalizada	Poisson Generalizada	30,86%
Magdalena	Poisson Generalizada	Poisson Generalizada	65,77%
Sucre	Poisson Generalizada	Poisson Generalizada	53,64%

Regional	Distribución Seleccionada (Afiliados Recuperados del COVID)	Distribución Seleccionada (Afiliados no Infectados por COVID)	Incremento en las utilizaciones
Cesar	Poisson Generalizada	Poisson Generalizada	53,22%

Tabla 31. Incremento Utilizaciones Esperadas 2023.

Se logra observar cómo existe un incremento notable en el número de utilizaciones esperadas para un afiliado que se ha recuperado del covid-19 respecto a aquellos que nunca contrajeron la enfermedad, este efecto se observa a lo largo de las distintas regionales siendo magdalena aquella que manifiesta el mismo en mayor medida.

Costo Esperado por Afiliado 2023

A continuación, se procede a estimar costo esperado por usuario en el año 2023 para los usuarios recuperados del covid-19 y aquellos que no fueron contagiados por dicha enfermedad mediante el ajuste de funciones de densidad de probabilidad.

Ajuste de funciones de densidad de probabilidad – Afiliados no infectados por covid-19

Regional Atlántico.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Atlántico es el siguiente:

Distribución de probabilidad	Kolmogorov-Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5268	61336.5823	13494292 *	13494292 *	13494336 *
Exponential	0.3502	30087.572	14221729	14221729	14221740
Gamma	0.1674	4341.4332	13698842	13698842	13698864
Inverse Gaussian	0.9284	152361.6494	17586742	17586742	17586765
Log Normal	0.0542 *	598.7202 *	13511712	13511712	13511734
Pareto one $Y>0$	0.4425	33190.6477	14838424	14838424	14838435
Weibull	0.0734	878.5832	13550616	13550616	13550639

Tabla 32. Criterios de Ajuste – $x_{i,t}^{(1)}$ Regional Atlántico.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Log Normal y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

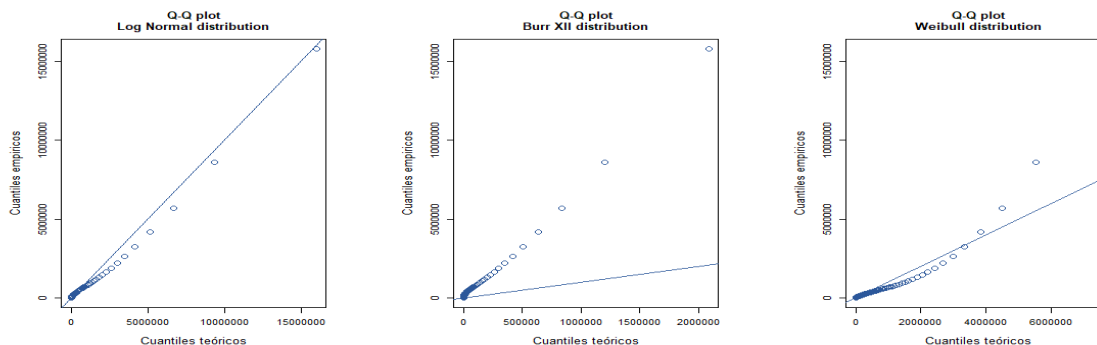


Figura 22. Gráficos Q-Q – $x_{i,t}^{(1)}$ Regional Atlántico.
Fuente. Elaboración Propia.

Se observa como la distribución Log Normal Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	11.987403	11.981806	11.993000
sigma	1.977485	1.973531	1.981447

Tabla 33. Parámetros del modelo seleccionado - Regional Atlántico.
Fuente. Elaboración Propia

Regional Bolívar.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Bolívar es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5156	91162.7079	19097438 *	19097438 *	19097484 *
Exponential	0.378	47635.4829	20348094	20348094	20348105
Gamma	0.1442	4731.4487	19344093	19344093	19344116
Inverse Gaussian	0.9716	223032.0258	26138224	26138224	26138247
Log Normal	0.0635	890.5952	19118069	19118069	19118092
Pareto one $Y>0$	0.4463	43633.3605	20808061	20808061	20808073
Weibull	0.0532 *	475.2196 *	19144386	19144386	19144409

Tabla 34. Criterios de Ajuste – $x_{14}^{(1)}$ Regional Bolívar.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Weibull y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

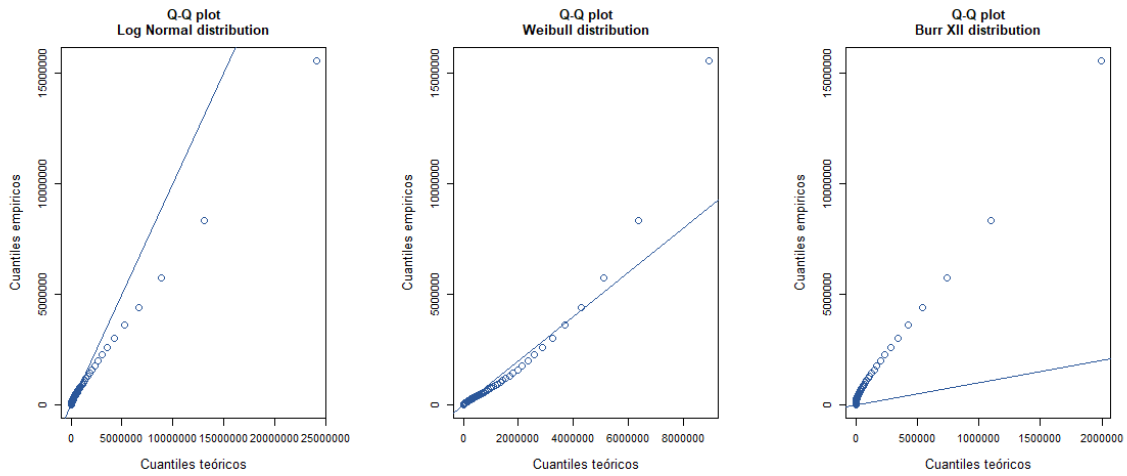


Figura 23. Gráficos Q-Q - $x_{i:n}^{(1)}$ Regional Bolívar.
Fuente. Elaboración Propia.

Se observa como la distribución Weibull Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	11.769025	11.763696	11.774354
sigma	2.247049	2.243284	2.250821

Tabla 35. Parámetros del modelo seleccionado - Regional Bolívar.
Fuente. Elaboración Propia.

Regional Córdoba.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Córdoba es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5308	69833.6325	14010149 *	14010149 *	14010193 *
Exponential	0.3028	23226.3975	14670919	14670919	14670930
Gamma	0.1148	2155.4296	14140215	14140215	14140237
Inverse Gaussian	0.9658	161653.8074	18970841	18970841	18970863
Log Normal	0.0745	937.3097	14037195	14037195	14037217
Pareto one $Y>0$	0.4435	32723.1265	15323571	15323571	15323582
Weibull	0.0406 *	243.0506 *	14031493	14031493	14031515

Tabla 36. Criterios de Ajuste – $x_{i,t}^{(1)}$ Regional Córdoba.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Weibull y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

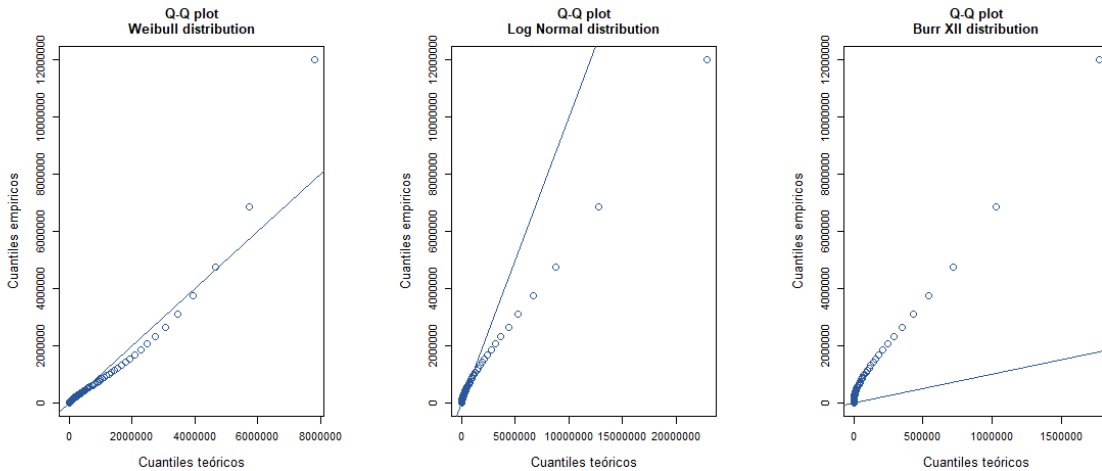


Figura 24. Gráficos Q-Q – $x_{i,t}^{(1)}$ Regional Córdoba.
Fuente. Elaboración Propia.

Se observa como la distribución Weibull Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	421979.987	419618.699	424354.563
sigma	0.52343012	0.52236957	0.52449282

Tabla 37. Parámetros del modelo seleccionado - Regional Córdoba.
Fuente. Elaboración Propia.

Regional Magdalena.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Magdalena es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5197	26334.3576	5663001 *	5663001 *	5663042 *
Exponential	0.3751	14019.4352	6010535	6010535	6010545
Gamma	0.1635	1808.6481	5753577	5753577	5753598
Inverse Gaussian	0.4685	17844.1056	5993769	5993769	5993790
Log Normal	0.0743	285.0638 *	5668323	5668323	5668343
Pareto one $Y>0$	0.4507	13754.159	6212238	6212238	6212248
Weibull	0.0628 *	292.5566	5684475	5684475	5684495

Tabla 38. Criterios de Ajuste – $x_{1,t}^{(1)}$ Regional Magdalena.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Log Normal y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

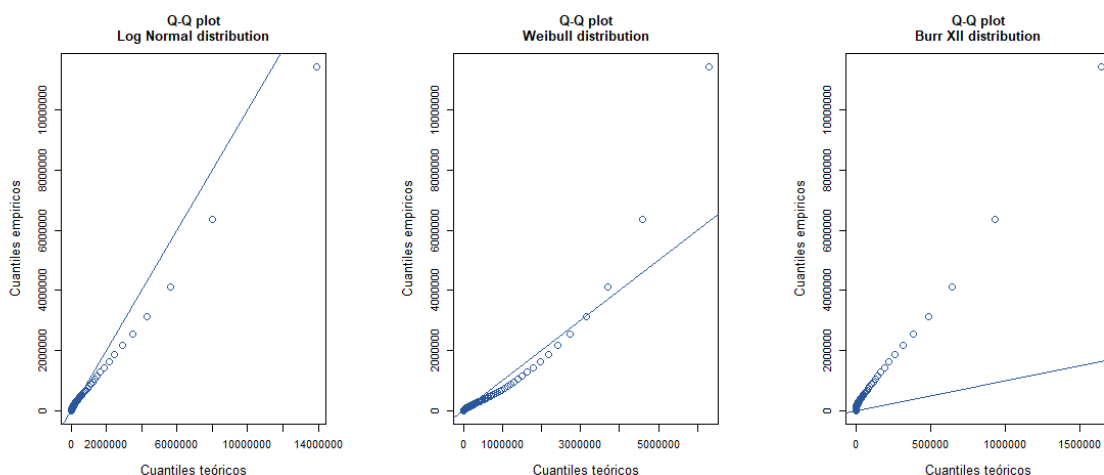


Figura 25. Gráficos Q-Q – $x_{i,t}^{(1)}$ Regional Magdalena.
Fuente. Elaboración Propia.

Se observa como la distribución Log Normal Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	11.698196	11.689358	11.707033
sigma	2.041118	2.034878	2.047376

Tabla 39. Criterios de Ajuste – $x_{i,t}^{(1)}$ Regional Magdalena.
Fuente. Elaboración Propia.

Regional Sucre.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Sucre es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5319	31575.4445	6548957 *	6548957 *	6548999 *
Exponential	0.3332	12511.01	6878001	6878001	6878011
Gamma	0.1351	1374.1788	6621056	6621056	6621077
Inverse Gaussian	0.5306	27139.3123	7040869	7040869	7040890
Log Normal	0.0718	377.1664	6560022	6560022	6560043
Pareto one $Y>0$	0.4396	15384.1979	7167833	7167833	7167843
Weibull	0.05 *	204.3081 *	6562954	6562954	6562974

Tabla 40. Criterios de Ajuste – $x_{i,t}^{(1)}$ Regional Sucre.

Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Weibull y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

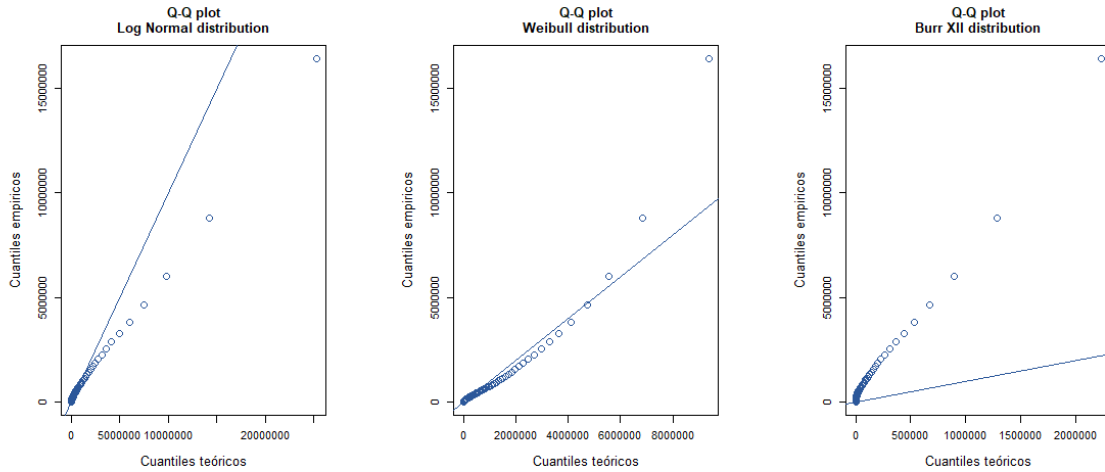


Figura 26. Gráficos Q-Q – $x_{i,t}^{(1)}$ Regional Sucre.
Fuente. Elaboración Propia.

Se observa como la distribución Weibull Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	328699.247	314966.677	343030.558
sigma	0.44928106	0.44336137	0.45527978

Tabla 41. Parámetros del modelo seleccionado - Regional Sucre.
Fuente. Elaboración Propia.

Regional Cesar.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Cesar es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.4977	1502.4786	323749 *	323749 *	323778 *
Exponential	0.4106	920.685	348854	348854	348862
Gamma	0.1366	78.1778	327847	327847	327862
Inverse Gaussian	0.5767	1802.2194	357438	357438	357453
Log Normal	0.0515 *	11.2809	323810	323810	323825
Pareto one $Y>0$	0.4513	708.9706	350802	350802	350810
Weibull	0.0689	6.1713 *	324574	324574	324588

Tabla 42. Criterios de Ajuste – $x_{it}^{(1)}$ Regional Cesar.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Weibull y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

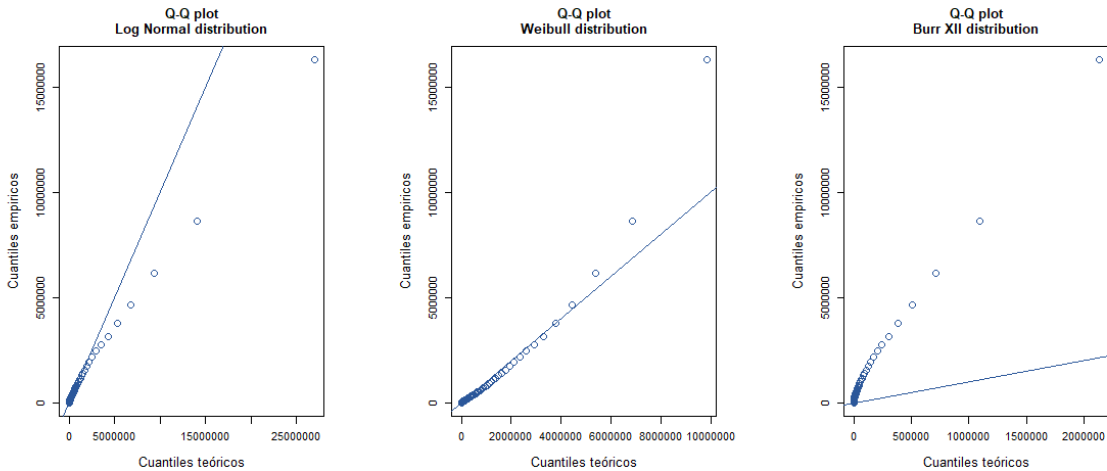


Figura 27. Gráficos Q-Q – $x_{it}^{(1)}$ Regional Cesar
Fuente. Elaboración Propia.

Se observa como la distribución Weibull Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	496904.996	492792.169	501052.149
sigma	0.52019242	0.51866206	0.52172729

Tabla 43. Parámetros del modelo seleccionado - Regional Cesar.
Fuente. Elaboración Propia.

Ajuste de funciones de densidad de probabilidad – Afiliados Recuperados del Covid-19

Regional Atlántico.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Atlántico es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5044	3773.4355	873483 *	873483 *	873516 *
Exponential	0.3979	2281.4668	926921	926921	926930
Gamma	0.1823	312.7213	887382	887382	887399
Inverse Gaussian	0.4944	3169.7167	930280	930280	930296
Log Normal	0.061 *	38.5174 *	874476	874476	874493
Pareto one $Y>0$	0.448	2068.7113	957125	957125	957133
Weibull	0.0795	59.7223	877120	877120	877136

Tabla 44. Criterios de Ajuste – $x_{i,t}^{(2)}$ Regional Atlántico.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Log Normal y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

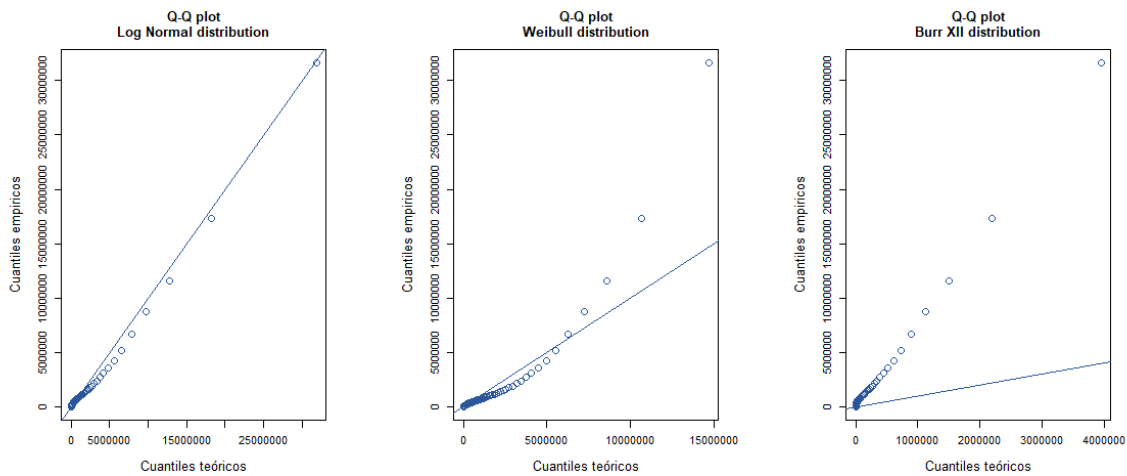


Figura 28. Gráficos Q-Q – $x_{it}^{(2)}$ Regional Atlántico.
Fuente. Elaboración Propia.

Se observa como la distribución Log Normal Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	12.480583	12.457220	12.503945
sigma	2.061166	2.044713	2.077753

Tabla 45. Parámetros del modelo seleccionado - Regional Atlántico.
Fuente. Elaboración Propia.

Regional Bolívar.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Bolívar es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5132	4158.7528	922707 *	922707 *	922740 *
Exponential	0.3665	2051.8243	974833	974833	974841
Gamma	0.1446	222.1971	933652	933652	933669
Inverse Gaussian	0.7739	8133.2102	1085994	1085994	1086010
Log Normal	0.0589	32.5379	923869	923869	923886
Pareto one $Y>0$	0.4407	2109.261	1007245	1007245	1007253

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Weibull	0.0529 *	28.7493 *	925128	925128	925145

Tabla 46. Criterios de Ajuste – $x_{it}^{(2)}$ Regional Bolívar.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Weibull y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

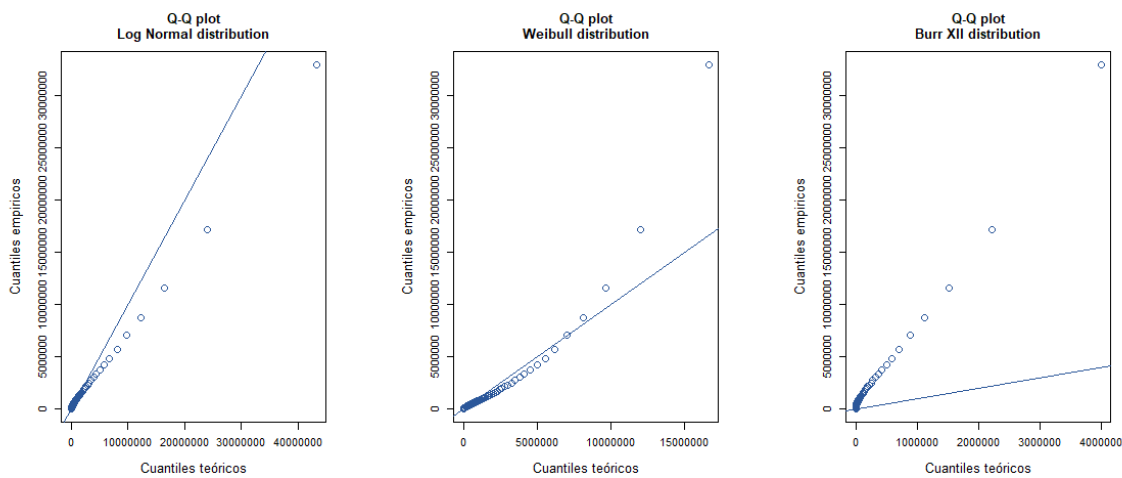


Figura 29. Gráficos Q-Q – $x_{it}^{(2)}$ Regional Bolívar.
Fuente. Elaboración Propia.

Se observa como la distribución Weibull Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	766921.201	749057.503	785210.917
sigma	0.4964705	0.49253579	0.50043664

Tabla 47. Parámetros del modelo seleccionado - Regional Bolívar.
Fuente. Elaboración Propia.

Regional Córdoba.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Córdoba es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5276	3737.1028	807711 *	807711 *	807744 *
Exponential	0.3382	1545.9972	847523	847523	847532
Gamma	0.1396	170.4031	815932	815932	815948
Inverse Gaussian	0.7107	5971.3821	921065	921065	921081
Log Normal	0.0704	43.4181	809302	809302	809319
Pareto one $Y>0$	0.437	1849.8286	882564	882564	882572
Weibull	0.0565 *	26.643 *	809371	809371	809387

Tabla 48. Criterios de Ajuste - $x_{i,t}^{(2)}$ Regional Córdoba.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Weibull y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

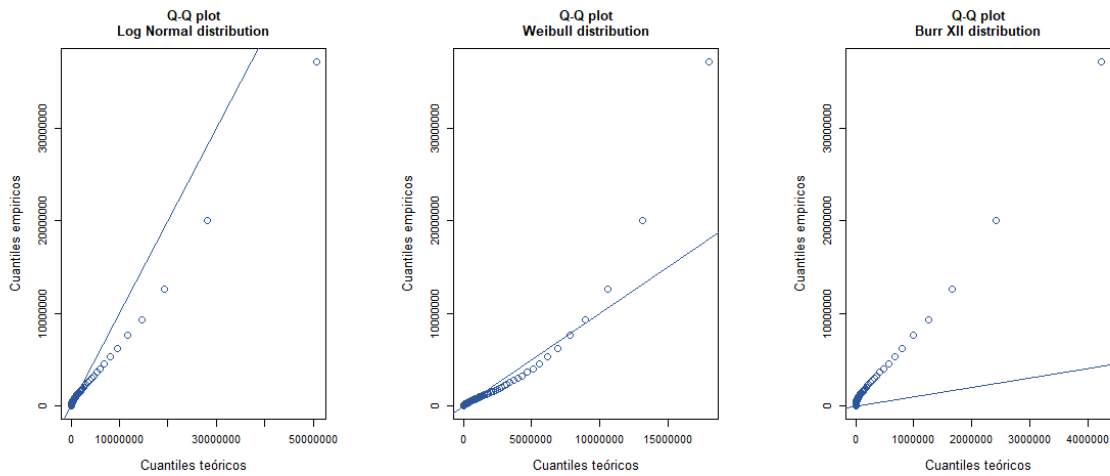


Figura 30. Gráficos Q-Q - $x_{i,t}^{(2)}$ Regional Córdoba.
Fuente. Elaboración Propia.

Se observa como la distribución Weibull Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	905450.059	883461.897	927985.477
sigma	0.51129106	0.5069175	0.51570234

Tabla 49. Parámetros del modelo seleccionado - Regional Córdoba.
Fuente. Elaboración Propia.

Regional Magdalena.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Magdalena es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5018	1409.9929	334337 *	334337 *	334367 *
Exponential	0.39	847.4601	354463	354463	354470
Gamma	0.1704	115.5097	339712	339712	339727
Inverse Gaussian	0.3902	734.2872	347823	347823	347837
Log Normal	0.0519 *	9.368 *	334562	334562	334577
Pareto one Y>0	0.4491	800.6085	366706	366706	366713
Weibull	0.0711	21.1527	335906	335906	335920

Tabla 50. Criterios de Ajuste – $x_{i,t}^{(2)}$ Regional Magdalena.
Fuente. Elaboración Propia.

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Log Normal y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

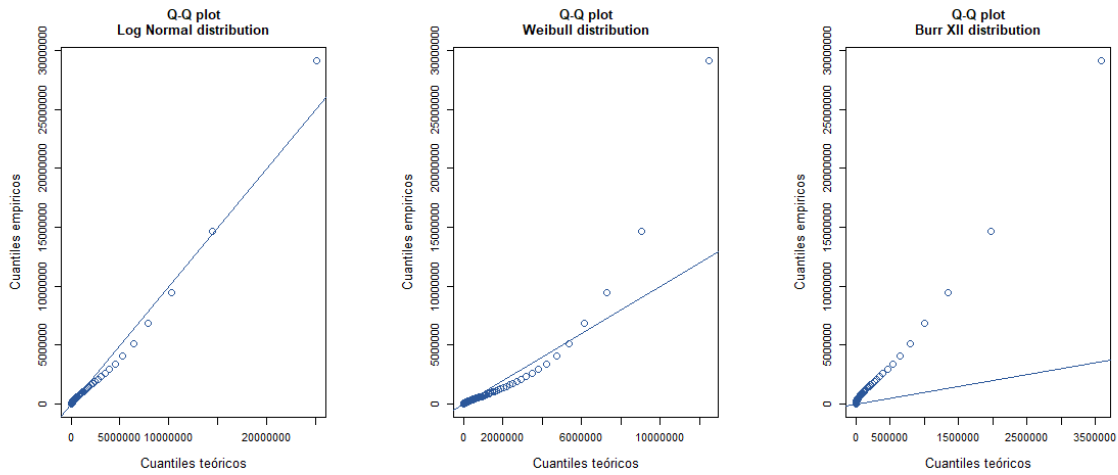


Figura 31. Gráficos Q-Q – $\chi_{ik}^{(2)}$ Regional Magdalena.
Fuente. Elaboración Propia.

Se observa como la distribución Log Normal Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	12.330899	12.294048	12.367750
sigma	2.022677	1.996787	2.048904

Tabla 51. Parámetros del modelo seleccionado - Regional Magdalena.
Fuente. Elaboración Propia.

Regional Sucre.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Sucre es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.5242	1442.3489	331516 *	331516 *	331545 *
Exponential	0.3826	781.8604	350409	350409	350416
Gamma	0.1609	100.226	336102	336102	336116
Inverse Gaussian	0.4767	1039.0892	349754	349754	349769
Log Normal	0.0654	12.7761 *	331900	331900	331915
Pareto one $Y>0$	0.4422	769.2272	362734	362734	362742
Weibull	0.0631 *	17.6634	332626	332626	332641

Tabla 52. Criterios de Ajuste – $x_{it}^{(2)}$ Regional Sucre.
Fuente. Elaboración Propia

Según los criterios de información, las distribuciones que presentan mejor ajuste son la Log Normal y la Burr XII.

Para seleccionar la mejor distribución se evalúan los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos.

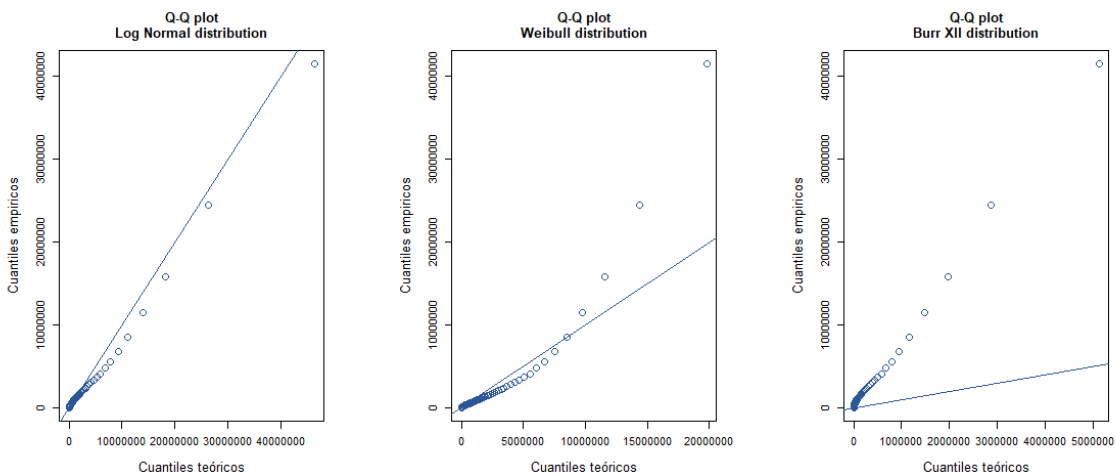


Figura 32. Gráficos Q-Q – $x_{it}^{(2)}$ Regional Sucre.
Fuente. Elaboración Propia.

Se observa como la distribución Log Normal Presenta un mejor ajuste en el Q-Q plot, por lo tanto, se selecciona esta como la mejor distribución.

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	12.767354	12.728349	12.80636
sigma	2.098297	2.070897	2.12606

Tabla 53. Parámetros del modelo seleccionado - Regional Sucre.
Fuente. Elaboración Propia

Regional Cesar.

El resumen sobre el ajuste de las distribuciones consideradas para la Regional Cesar es el siguiente:

Distribución de probabilidad	Kolmogorov -Smirnov	Cramer-von Mises	AIC	AICC	BIC
Burr XII	0.4754	69.1554	16318	16318	16335
Exponential	0.4421	51.8603	17714	17714	17718
Gamma	0.1519	4.613	16539	16539	16547
Inverse Gaussian	0.3314	22.7014	16656	16656	16665
Log Normal	0.0524 *	0.4067	16309 *	16309 *	16317 *
Pareto one $Y>0$	0.4486	34.9204	17639	17639	17643
Weibull	0.0724	0.3819 *	16367	16367	16376

Tabla 54. Criterios de Ajuste – $x_{it}^{(2)}$ Regional Cesar.
Fuente. Elaboración Propia.

Según los criterios de información, la distribución que presenta mejor ajuste es la Log Normal

Asimismo, al comparar los gráficos Q-Q de las tres mejores distribuciones frente a los percentiles empíricos se observa como la distribución Log Normal presenta un mejor ajuste.

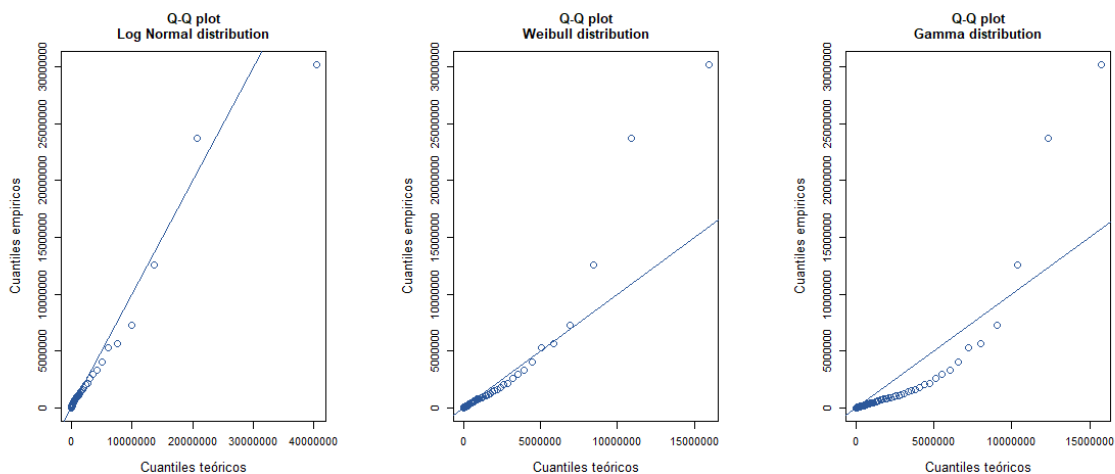


Figura 33. Gráficos Q-Q – $x_{i,k}^{(2)}$ Regional Cesar.
Fuente. Elaboración Propia.

Dicho lo anterior se selecciona a la distribución Log Normal como el mejor modelo

Los parámetros estimados del modelo seleccionado son:

Parámetro	Estimación	2.5 %	97.5 %
mu	11.813780	11.613866	12.013694
sigma	2.450096	2.312736	2.595614

Tabla 55. Parámetros del modelo seleccionado - Regional Bolívar.
Fuente. Elaboración Propia.

Resumen - Costo Esperado por Afiliado 2023 (Usuarios Recuperados del Covid Vs No Infectados)

Una vez se han ajustado las distribuciones de probabilidad, se procede a calcular el incremento en el costo esperado por afiliado para el año 2023. La Tabla 56 presenta los incrementos de las utilidades esperadas de los usuarios recuperados por covid frente a aquellos que nunca fueron infectados por dicho virus.

Regional	Distribución Seleccionada (Afiliados Recuperados del Covid)	Distribución Seleccionada (Afiliados no Infectados por Covid)	Incremento esperado en la severidad 2023
Atlántico	Log Normal	Log Normal	95,77%
Bolívar	Weibull	Weibull	85,21%
Córdoba	Weibull	Weibull	57,26%
Magdalena	Log Normal	Log Normal	129,10%
Sucre	Log Normal	Weibull	91,31%
Cesar	Log Normal	Weibull	119,86%

Tabla 56. Incremento Costo Esperado por Afiliado 2023.

Capítulo III – Conclusiones

Actualmente, los sistemas sanitarios y las compañías aseguradoras enfrentan un nuevo reto en relación con la atención de aquellos pacientes que se han recuperado del COVID-19. En el presente estudio se ha comprobado que estos individuos presentan riesgos de salud latentes y, en consecuencia, los costos asociados a su atención médica son más elevados en comparación con aquellos que no contrajeron la enfermedad, y así mismo quedó demostrado que cada una de las regionales que participaron en este estudio poseen un mayor o menor impacto por parte del llamado “Long Covid”.

Así mismo se logró observar que pese a que Colombia y concretamente los departamentos presentes en este estudio ostenten más de un 80% de dosis aplicadas de su población con esquema completo de vacunación (de acuerdo con cifras del ministerio de salud) tanto el costo medio de atención de un paciente en etapa aguda del covid como el número de utilidades de servicios de salud no presentan una disminución.

La pandemia de COVID-19 ha representado, sin lugar a duda, un desafío sin precedentes para la solidez y resiliencia de los sistemas sanitarios a nivel global. Este fenómeno, más que un simple episodio en la historia de la humanidad ha dejado huellas indelebles en una parte significativa de la población. Al concluir el estudio en cuestión, se pudo determinar el impacto duradero que ha tenido el virus en la vida de las personas.

Ante esta situación, es crucial implementar estrategias de detección temprana y llevar a cabo actividades enfocadas en mejorar la salud y calidad de vida de la población afiliada. Esto incluye la promoción de hábitos saludables, prevención de enfermedades y la atención oportuna y adecuada a las necesidades específicas de cada individuo.

Esto último debe ser un compromiso por parte de los sistemas sanitarios, los cuales a través de los distintos mecanismos que posean para la gestión del riesgo en salud deben incluir a aquellas personas recuperadas del covid-19 como un grupo de principal atención, tanto por la connotación social que esto representa para con la población como la arista financiera que implica una población con riesgos en salud no intervenidos previamente.

Capítulo IV – Trabajos Futuros.

La naturaleza descriptiva del presente proyecto fija un punto de partida en materia de investigación en cuanto a los efectos a largo plazo del covid-19. Como ya se ha presentado, los individuos que se recuperaron de dicha enfermedad poseen un mayor costo esperado y una mayor severidad respecto a aquellos que nunca contrajeron la enfermedad.

Con este trabajo como precedente se justifican estudios mas específicos de índole inferencial los cuales puedan entrarse en las distintas variables relacionadas a cada uno de los Individuos y su incidencia en aspectos tales como la frecuencia, severidad y salud individual.

Adicionalmente, cobra relevancia la realización de estudios longitudinales para evaluar la persistencia de la mayor frecuencia y severidad observadas en los afectados a largo plazo. Además, dado que el presente estudio se concentra únicamente en 6 departamentos, es crucial ampliar la muestra de estudio, incluyendo una diversidad de regiones y demografías, para comprender mejor la generalización de estos resultados. Un enfoque significativo sería la segmentación de pacientes recuperados de COVID-19 basada en variables como edad, género, comorbilidades y severidad de la infección inicial, lo que permitiría identificar subgrupos con perfiles de riesgo diferenciados.

Asimismo, dada la emergencia de variantes del virus SARS-CoV-2, es necesario analizar el impacto específico de cada variante en los resultados de salud. La investigación de factores de riesgo contributivos, tales como edad, comorbilidades y hábitos de vida, podría proporcionar información valiosa sobre qué grupos son más susceptibles a consecuencias graves.

Un aspecto de vital importancia corresponde a la evaluación del impacto de las vacunas contra el COVID-19 en la moderación de la frecuencia y severidad de problemas de salud post-infección.

Desde un enfoque actuarial, el desarrollo de modelos más avanzados es imperativo para una evaluación de riesgos y costos más precisa. Tal y como se observó en el presente estudio, las distribuciones seleccionadas para modelar la severidad pertenecen a la familia exponencial, lo cual permitiría el modelamiento de la misma mediante metodologías tales como los modelos lineales generalizados, los cuales permiten incluir variables exógenas las cuales pueden brindar un mayor entendimiento de la severidad.

A nivel de seguros de salud, es relevante explorar cómo estos hallazgos pueden influir en las políticas y estructuras de primas. Finalmente, realizar comparaciones con otras enfermedades virales puede ofrecer una perspectiva sobre si el COVID-19 posee un perfil de riesgo distintivo. Al abordar estos aspectos en futuras investigaciones, podemos contribuir a una comprensión más completa y matizada de las implicaciones a largo plazo del COVID-19 en la salud de las personas y la sociedad en general.

Referencias

- Andersen, R. M. (1995). Revisiting the Behavioral Model and Access to Medical Care: Does it Matter? *Journal of Health and Social Behavior*.
- Brodin P. (2021). Immune determinants of COVID-19 disease presentation and severity. *Nat Med*, 27: 28–33.
- Callard And Perego, B. (2021). *How and why patients made Long Covid*.
- Charpentier, M. &. (2005). *MATHÉMATIQUES DE L'ASSURANCE NON-VIE*. Économie et Statistiques Avancées.
- Cochran, W. G. (1977). *Sampling techniques. Third Edition*.
- Davido B, S. S. (2020). Post-COVID-19 chronic symptoms: a postinfectious. *Clin Microbiol Infect*, 26: 1448–1449.
- FAIR Health. (2022). *Patients Diagnosed with Post-COVID* .
- Fernández de Las Peñas C. (2021). Defining post-COVID symptoms (post-acute COVID, long COVID, persistent post-COVID): an integrative classification. *Int J Environ Res Public Health*, 18: 2621.
- Folland, S. G. (2017). *The Economics of Health and Health Care: International Student Edition*.
- George Casella, R. L. (2002). *Statistical Inference 2nd Edition*.
- Gerdtham and Jönsson, B. (2000). Determinants of Health Expenditures: Stylized Facts and a New Signal.
- Juan-Manuel Anaya, e. a. (2021). *Post-COVID syndrome. A case series and comprehensive review*.
- Klugman, S. A. (2019). *Loss Models From Data Decisions*.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*.
- McDonald, J. B. (1984). *Some Generalized Functions for the Size Distribution of Income*.
- Mikosch, T. (2009). *Non-life insurance mathematics: an introduction with the Poisson process*. Berlin: Springer.
- Mikosch, T. (2014). *Non-Life Insurance Mathematics*. Springer, 2014.
- Ministerio de Sanidad de España. (2021). *Enfermedad por coronavirus, COVID-19*.
- MinSalud. (2018). *Gestión Integral Del Riesgo En salud*. Bogotá.

- MinSaud. (2022). *PLAN DECENAL DE SALUD PÚBLICA 2022 - 2031*.
- Munipalli, B. S. (2022). Post-acute sequelae of COVID-19 (PASC): a meta-narrative review of pathophysiology, prevalence, and management.
- Nalbandian A, S. K. (2021). Post-acute sequelae of SARS-CoV-2 infection and subjective memory problems. *JAMA Netw Open*, 4: e2119335.
- Nath A. (2020). Long-Haul COVID. . *Neurology*, 95: 559–560.
- Norman L. Johnson, S. K. (1995). *Continuous Univariate Distributions, Vol. 2*.
- Oh ES, V. T. (2021). Post-acute sequelae of SARS-CoV-2 infection and subjective memory problems. *JAMA Netw Open*, 4: e2119335.
- OPS. (2021, junio). Retrieved from https://iris.paho.org/bitstream/handle/10665.2/55389/OPSNMHHMHCVID-19210018_spa.pdf?sequence=1&isAllowed=y
- Phillips S and Williams MA. (2021). Confronting our next national health disaster Long-Haul Covid. *N Engl J Med*, 385: 577–579.
- Robert Rigby, M. S. (2017). *Distributions for Modelling Location, Scale And Shape Using GAMLSS in R*.
- Shah W, H. T. (2021). Managing the long term effects of covid-19: summary of NICE, SIGN, and RCGP rapid guideline. *BMJ*, 372: n136.
- Sivan M and Taylor S. (2020). NICE guideline on long COVID. *BMJ*, 371: m4938.
- SJ, Y. (2021). Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. *Infect Dis*, 53: 737–754.
- Stefanou MI, P. L. (2022). Neurological manifestations of long-COVID syndrome: a narrative review.
- Thomas Golaszewski, W. L. (1989). *The Relationship between Retrospective Health Insurance Claims and a Health Risk Appraisal-Generated Measure of Health Status*.
- Wagstaff, A., & Claeson, M. (2004). The Millennium Development Goals for Health : Rising to the Challenges.
- Weerahandi, H. H. (2021). *Post-Discharge Health Status and Symptoms in Patients with Severe COVID-19*.

World Health Organization. (2023). Retrieved from https://www.who.int/health-topics/coronavirus#tab=tab_1

World Health Organization, W. (2020). *Origin of SARS-CoV-2*. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/332197/WHO-2019-nCoV-FAQ-Virus_origin-2020.1-eng.pdf

Ziyad Al-Aly, Y. X. (2021). High-dimensional characterization of post-acute sequelae of COVID-19.

Anexos.

Anexo 1. Encuesta Aplicada.

Pregunta.
¿Cuál es su situación laboral?
¿Cuál es su nivel educativo?
¿A que estrato socioeconómico pertenece?
¿En que tipo de zona reside actualmente?
¿Una vez recuperado del covid-19 ha sufrido usted de fatiga?
¿Una vez recuperado del covid-19 ha presentado dificultades para dormir?
¿Una vez recuperado del covid-19 ha presentado pérdida parcial o total del sentido del olfato?
¿Una vez recuperado del covid-19 ha sufrido usted de Dolor de cabeza frecuente?
¿Una vez recuperado del covid-19 ha sufrido usted de dolor en el pecho frecuente?
¿Una vez recuperado del covid-19 ha sufrido usted de problemas estomacales frecuentes?
¿Una vez recuperado del covid-19 ha sufrido usted de dificultad respiratoria?
¿Ha presentado algún síntoma adicional a los antes mencionados?
¿Cual?
Una vez recuperado del covid-19, ¿Es usted capaz de caminar?
Una vez recuperado del covid-19, ¿Fue usted capaz de retomar sus actividades cotidianas (Trabajo, Estudio, Ejercicio, y demás actividades)?
Una vez recuperado del covid-19, ¿Ha sufrido de algún tipo de dolencias?
¿Una vez recuperado del covid-19 ha sufrido Depresión?
¿Una vez recuperado del covid-19 ha sufrido Ansiedad?

Anexo 2. Código en R

```
# 1. Set Up -----
```

```
library(lubridate)
```

```
library(tidyverse)
```

```
library(DBI)
```

```
library(stringr)
```

```
library(bigrquery)
```

```
library(scales)
```

```
library(readxl)
```

```
library(CreateBQTables)
```

```
library(lemon)
```

```
options(scipen = 999999)
```

```
`%!in%` = Negate(`%in%`)
```

```
# Tabla Agregada de Siniestros
```

```
Siniestros <- Siniestros
```

```
# Maestros de Afiliados
```

```
Eafil <- Eafil
```

```
# GetFixedQuery(Eafil %>%
```

```
# mutate_at(vars(AFIN_GRUPO_POBLACIONAL, AFIN_UBICACION_GEOGRAFICA,  
AFIN_NIVEL_SISBEN, AFIN_METODOLOGIA_SISBEN, AFIN_GRUPO_ETARIO,
```

```
#           CONTRATOPRIMARIO, AFIN_ALFANET), as.numeric), CopyToClipboard = T)
```

```

# Base de Siniestros (Afiliados con Enfermedades de alto costo)

Esin <- tbl(con, "REPORTE_ESIN_SINIESTROS")

# Base de Casos Positivos Covid

PositivosCovid <- tbl(con, "Detalle_Positivos_Covid")

# Se calcula el numero de afiliados unicos con resultado positivo para covid

NumPositivos <- PositivosCovid %>% summarise(Result = n_distinct(IDEN_AFIL)) %>% collect()
%>% pull()

Inflacion <- read_excel("Data/Inflacion.xlsx")

# Funcion que Genera la grafica de piramides poblacionales.

PlotPiramide <- function(Agregado_Caracterizacion, GruposEtarios, Genero, NumPersonas, Titulo){

  Datos_Piramide <- Agregado_Caracterizacion %>%

  dplyr::select(GruposEtarios, Genero, NumPersonas) %>%

  data.table::setnames(old = c(GruposEtarios, Genero, NumPersonas), new = c("GruposEtarios",
"Genero", "NumPersonas")) %>%

  group_by(GruposEtarios, Genero) %>%

```

```

summarise(n=sum(NumPersonas)) %>%
ungroup() %>%
#group_by(Genero) %>%
mutate(FREQ=round(100*n/sum(n),1))

limite_c <- floor(max(Datos_Piramide$n) + .2*max(Datos_Piramide$n))

Piramide <- ggplot(Datos_Piramide,
  aes(x=ifelse(test = Genero == 'M', yes = -n, no = n),
    y=GruposEtarios,
    fill=factor(Genero, levels=c('M','F')))) +
geom_col() +
geom_text(aes(label=paste0(FREQ,'%')), size=3,
  hjust=ifelse(test = Datos_Piramide$Genero == 'M', yes = 1.1, no = -0.1)) +
scale_x_symmetric(labels = abs) +
labs(x = 'Número de afiliados', y='Grupo etario', fill='') +
scale_fill_manual(labels=c('Hombres','Mujeres'), values=c('#505050','#d23f67')) +
coord_cartesian(xlim = c(-limite_c,limite_c)) +
ggtitle(Titulo) +
theme_classic() +
theme(legend.position = 'bottom', legend.box = 'horizontal',
  strip.text.x = element_text(size = 15),
  panel.spacing = unit(1.5, "lines"),

```



```

    plot.title = element_text(color="black", size=14, face="bold.italic", hjust = 0.5))
  return(list(Piramide, Datos_Piramide))
}

```

```

PlotPiramide_2 <- function(Agregado_Caracterizacion, GruposEtarios, Genero, NumPersonas,
Titulo, Segmentos, nrow=2){

```

```

  Datos_Piramide <- Agregado_Caracterizacion %>%

```

```

  dplyr::select(GruposEtarios, Genero, NumPersonas, Segmentos) %>%

```

```

  data.table::setnames(old = c(GruposEtarios, Genero, NumPersonas, Segmentos),

```

```

    new = c("GruposEtarios", "Genero", "NumPersonas", "Segmentos")) %>%

```

```

  group_by(GruposEtarios, Genero, Segmentos) %>%

```

```

  summarise(n=sum(NumPersonas)) %>%

```

```

  ungroup() %>%

```

```

  group_by(Segmentos) %>%

```

```

  mutate(FREQ=round(100*n/sum(n),1))

```

```

  limite_c <- floor(max(Datos_Piramide$n) + .2*max(Datos_Piramide$n))

```

```

  Piramide <- ggplot(Datos_Piramide,

```

```

    aes(x=ifelse(test = Genero == 'M', yes = -n, no = n),

```

```

      y=GruposEtarios,
      fill=factor(Genero, levels=c('M','F')))) +
geom_col() +
geom_text(aes(label=paste0(FREQ,'%')), size=3,
          hjust=ifelse(test = Datos_Piramide$Genero == 'M', yes = 1.1, no = -0.1)) +
scale_x_symmetric(labels = abs) +
labs(x='Número de afiliados', y='Grupo etario', fill='') +
scale_fill_manual(labels=c('Hombres','Mujeres'), values=c('#505050','#d23f67')) +
coord_cartesian(xlim = c(-limite_c,limite_c)) +
facet_wrap(~Segmentos, nrow = nrow) +
ggtitle(Titulo) +
theme_classic() +
theme(legend.position = 'bottom', legend.box = 'horizontal',
      strip.text.x = element_text(size = 15),
      panel.spacing = unit(1.5, "lines"),
      plot.title = element_text(color="black", size=14, face="bold.italic", hjust = 0.5))
return(list(Piramide, Datos_Piramide))
}

# Duncion para copiar resultados al portapapeles
CopyClipboard <- function(x, col.names = TRUE, row.names=FALSE){
  write.table(x, "clipboard", col.names = col.names, row.names=row.names, sep="\t", dec = ",")
}

```

```
Hospitalizaciones <- tbl(con, "REPORTE_RIPS_HOSPITALARIOS")
```

```
Autorizaciones <- tbl(con, "REPORTE_AUTORIZACIONES")
```

```
# 2. Preparacion de las bases de datos -----
```

```
## 2.1 Preparacion de Bases Auxiliares -----
```

```
# Se genera una tabla de afiliados unicos para cada mes de reporte
```

```
# Con el fin de identificar su estado y las fechas de cambio de estado
```

```
# a traves del tiempo.
```

```
Distinct_Afil <- Eafil %>%
```

```
  mutate(IDEN_AFIL = sql("CONCAT(AFIC_TIPO_DOCUMENTO, ' - ', AFIC_DOCUMENTO)")) %>%
```

```
  mutate(reporte_mes = sql("DATE_TRUNC(FECHA_CORTE, MONTH)")) %>%
```

```
  distinct(IDEN_AFIL, reporte_mes, AFIC_ESTADO_AFILIADO, AFIF_RETIRO)
```

```
Agregado_Afil_Primer_Ult_est <- Eafil %>%
```

```
# Se genera un agregado de los afiliados para identificar su permanencia.
```

```
  mutate(IDEN_AFIL = sql("CONCAT(AFIC_TIPO_DOCUMENTO, ' - ', AFIC_DOCUMENTO)")) %>%
```

```

mutate(reporte_mes = sql("DATE_TRUNC(FECHA_CORTE, MONTH)")) %>%
group_by(IDEN_AFIL) %>%
summarise(Fecha_Inicial = min(reporte_mes),
          Fecha_Final = max(reporte_mes),
          min_Fecha_Novedad = min(AFIF_RETIRO),
          max_Fecha_Novedad = max(AFIF_RETIRO),
          Fecha_Nacimiento = max(AFIF_NACIMIENTO)) %>%
ungroup() %>%

```

```

# Se busca el Estado_Inicial del afiliado, para ello se cruza
# la primera fecha en la que aparecio el afiliado. Es decir,
# Si el afiliado aparecio por primera vez en julio de 2020, se
# Buscara que estado tenia en ese mes.

```

```

left_join(Distinct_Afil %>% dplyr::select(-c(AFIF_RETIRO)),
          by =c("IDEN_AFIL"="IDEN_AFIL",
               "Fecha_Inicial"="reporte_mes")) %>%

```

```

# Se busca el Estado_Final del afiliado, para ello se cruza
# la ultima fecha en la que aparecio el afiliado. Es decir,
# Si el afiliado aparecio por ultima vez en julio de 2020, se
# Buscara que estado tenia en ese mes.

```

```

left_join(Distinct_Afil %>% dplyr::select(-c(AFIF_RETIRO)),
  by =c("IDEN_AFIL"="IDEN_AFIL",
    "Fecha_Final"="reporte_mes"),
  suffix = c("_Inicial", "_Final"))

```

```

Marcacion_Cohortes <- Esin %>%
  filter(!is.na(REGIONAL)) %>%
  mutate(IDEN_AFIL = sql("CONCAT(TIPO_DOCUMENTO, ' - ',DOCUMENTO )")) %>%
  group_by(IDEN_AFIL, REGIONAL) %>%
  summarise(Recuento_Cohortes = n_distinct(CONC_NOMBRE)) %>%
  ungroup() %>%
  mutate(Segmentacion_num_Cohortes = case_when(
    Recuento_Cohortes == 1 ~ "1",
    Recuento_Cohortes == 2 ~ "2",
    Recuento_Cohortes == 3 ~ "3",
    Recuento_Cohortes >= 4 ~ "4+"
  )) %>%
  dplyr::select(-c(REGIONAL))

```

```

DetalleHops <- Hospitalizaciones %>%

```

Se crea un identificador de usuarios para poder usar la funcion n_distinc

```
mutate(IDEN_AFIL = sql("CONCAT(FCDC_TIPO_DOC_AFILIADO, ' - ', FCDC_AFILIADO)") %>%
```

Se selecciona los registros correspondientes al archivo AH el cual contiene el detalle de hospitalizaciones

```
filter(FCDC_ARCHIVO == "AH") %>%
```

Se convierten las siguientes variables a caracter para realizar cruces de informacion

```
mutate_at(vars(FCDV_PROVEEDOR, FCDN_CONTRATO), as.character) %>%
```

Se calculan los dias de hospitalizacion

```
mutate(Dias_Hosp = sql("DATE_DIFF(FECHA_EGRESO, FECHA_INGRESO, DAY)") %>%
```

Se eliminan Duplicados

```
mutate(fila_d=sql("ROW_NUMBER() OVER (PARTITION BY FCDC_FACTURA, FCDV_PROVEEDOR,  
FCDN_CONTRATO,FECHA_INGRESO, IDEN_AFIL)") %>%
```

```
filter(fila_d==1) %>%
```

```
ungroup() %>% dplyr::select(-c(fila_d)) %>%
```

Se extrae el año de inicio de la hospitalizacion, este sera utilizado para escoger las hospitalizaciones a tener en cuenta

```
mutate(FECHA_INGRESO_ANHO = sql("EXTRACT(YEAR FROM FECHA_INGRESO)") %>%
```

Se agrega el codigo de servicio de la autorizacion, dado que los rips no lo reportan correctamente.

```

left_join(

# Se crea el listado de codigos de servicio por autorizacion.

Maestro_Cod_Serv <- Autorizaciones %>%

# Se excluyen errores

filter(IDAUTORIZACION != 0) %>%

# Se seleccionan las autorizaciones a tener en cuenta

filter(ESTADO == "NU") %>%

# Al seleccionar solo las autorizaciones de ambito A

# No aparecen hosp de media complejidad en rips.

# Al validar algunos casos se observa que se facturan hospitalizaciones

# con autorizaciones de ambito ambulatorio.

# llama la atencion que aunque a priori el servicio no deberia desembocar en una autorizacion.

# existen autorizaciones reportadas por ese servicio.

# Se deben validar los casos.

filter(AMBITO == "H") %>%

filter(CLASE == "PBS") %>%

group_by(IDAUTORIZACION, IDPRESTADOR, COD_SERVICIO) %>%

summarise(n=n()) %>%

ungroup() %>%

```

```

mutate(IDPRESTADOR = as.character(IDPRESTADOR)),

by = c("FCDN_AUTORIZACION"="IDAUTORIZACION",
      "FCDV_PROVEEDOR"="IDPRESTADOR")) %>%
mutate(Cruce = if_else(is.na(COD_SERVICIO), 0, 1)) %>%
filter(!is.na(FCDN_AUTORIZACION)) %>%
filter(FCDN_AUTORIZACION != 0) %>%
filter(!is.na(FECHA_INGRESO)) %>%
group_by(IDEN_AFIL, FCDV_PROVEEDOR, FCDN_AUTORIZACION) %>%
summarise(FechaMinIngresoHops = min(FECHA_INGRESO),
          FechaMaxEgresoHosp = max(FECHA_EGRESO)) %>%
ungroup() %>%
mutate(Dias_Hosp = sql("DATE_DIFF(FechaMaxEgresoHosp, FechaMinIngresoHops, DAY)")) %>%
filter(Dias_Hosp>=1)

```

2.2 Preparacion Base Siniestros Covid -----

Se valida que no se hayan duplicado registros al momento de hacer el cruce, para ello se cuenta el numero de registros

De la tabla principal (Siniestros) antes y después de hacer el cruce sin aplicar ningún tipo de filtros. Los resultados fueron:

Registros Antes del cruce: 294302453

Registros Después del cruce: 294302453

```
Names_MarcasEsin <- Siniestros %>%
```

```
  dplyr::select(starts_with("ESIN_")) %>%
```

```
  colnames()
```

```
NamesRelevantes <- Siniestros %>%
```

```
  dplyr::select(1:44, PBS_NOPBS_2022_, marca_cambio_PBS_NOPBS) %>%
```

```
  colnames()
```

```
FinalNames <- c(NamesRelevantes, Names_MarcasEsin)
```

```
Afil_CalVida <- BaseEncuesta %>%
```

```
  filter(Final == "Manifiesta", BloquePreguntas == "Calidad de Vida") %>%
```

```
  distinct(IDEN_AFIL) %>%
```

```
  pull()
```

```
Afil_Postcovid <- BaseEncuesta %>%
```

```
  filter(Final == "Manifiesta", BloquePreguntas == "Post Covid") %>%
```

```
  distinct(IDEN_AFIL) %>%
```

```
  pull()
```

```
SiniestrosCovidProcesada <- Siniestros %>%
```

```
  dplyr::select(all_of(FinalNames)) %>%
```

```
  mutate(IDEN_AFIL = sql("CONCAT(TIPOIDUSUARIO, ' - ', IDUSUARIO)")) %>%
```

```
  mutate(ANHO_PRES = sql("EXTRACT(YEAR FROM FPRES)")) %>%
```

```
  mutate(MES_PRES = sql("EXTRACT(MONTH FROM FPRES)")) %>%
```

```
# Se marcan los siniestros de covid
```

```
left_join(
```

```
  PositivosCovid %>%
```

```
# Se seleccionan las variables relevantes
```

```
  dplyr::select(IDEN_AFIL, Fecha_Entrega_Resultado_min, Fecha_Entrega_Resultado_max,  
fallecido, nombre_dpto_atencion) %>%
```

```

# Se deja un unico registro por persona

mutate(fila_d = sql('ROW_NUMBER() OVER (PARTITION BY IDEN_AFIL)') %>% filter(fila_d==1)
%>%

dplyr::select(-c(fila_d)) %>%

ungroup() %>%

# Se Cuenta el numero de pruebas positivas por afiliado.

left_join(

PositivosCovid %>%

group_by(IDEN_AFIL) %>%

summarise(Numero_Pruebas_Positivas = n_distinct(id_toma_muestra_zscovid)) %>%

ungroup(), by = "IDEN_AFIL"

) %>%

# Se agrega una variable indicadora para marcar a los usuarios con diagnostico positivo

mutate(Afiliado_Covid = 1), by = "IDEN_AFIL") %>%

# Se reemplazan los "NA" de la variable Afiliado_Covid por ceros.

mutate(Afiliado_Covid = if_else(is.na(Afiliado_Covid), 0, Afiliado_Covid)) %>%

# Se seleccionan solamente aquellos afiliados que tienen un resultado positivo de covid.

#filter(Afiliado_Covid == 1) %>%

```

Se genera una nueva fecha la cual corresponde al periodo en el que usualmente el paciente se ha recuperado del covid.

```
mutate(Fecha_Entrega_Resultado_max_Mas6semanas =  
sql("DATE_ADD(Fecha_Entrega_Resultado_max, INTERVAL 6 WEEK)")) %>%
```

Se clasifican los siniestros en la base de datos

```
mutate(Tipo_Siniestro = case_when(  
  
  FPRES < Fecha_Entrega_Resultado_min ~ 0,  
  
  between(x = FPRES, left = Fecha_Entrega_Resultado_min,  
Fecha_Entrega_Resultado_max_Mas6semanas) ~ 1,  
  
  FPRES > Fecha_Entrega_Resultado_max_Mas6semanas ~ 2)) %>%
```

```
mutate(Des_Tipo_Siniestro = case_when(  
  
  Tipo_Siniestro == 0 ~ "PRE COVID",  
  
  Tipo_Siniestro == 1 ~ "DURANTE COVID",  
  
  Tipo_Siniestro == 2 ~ "POST COVID",  
  
  TRUE~NA_character_)) %>%
```

Se Busca la informacion relacionada a cohortes de a alto costo (COHORTE, FECHA DE DIAGNOSTICO)

```
left_join(  
  
  Esin %>%  
  
  filter(CONC_NOMBRE != "UCI") %>%
```

```

mutate(IDEN_AFIL = sql("CONCAT(TIPO_DOCUMENTO, ' - ', DOCUMENTO)")) %>%

# Se seleccionan los campos relevantes

dplyr::select(IDEN_AFIL, SINF_FECHA_SINIESTRO, CONC_NOMBRE) %>%

# Se deja un unico registro por persona (Una sola Cohorte)

mutate(fila_d = sql('ROW_NUMBER() OVER (PARTITION BY IDEN_AFIL)')) %>% filter(fila_d==1)
%>% dplyr::select(-c(fila_d)) %>% ungroup() %>% #Distinct con .keep_all = T

mutate(Esin = 1) %>%

mutate(SINF_FECHA_SINIESTRO = sql("PARSE_DATE('%d/%m/%Y', SINF_FECHA_SINIESTRO)")),
by = "IDEN_AFIL") %>%

# Se reemplazan los "NA" de la variable Esin por ceros.

mutate(Esin = if_else(is.na(Esin), 0, Esin)) %>%

mutate(Esin_Postcovid = if_else(SINF_FECHA_SINIESTRO > Fecha_Entrega_Resultado_min, 1, 0))
%>%

# Se añade el numero de cohortes a las que pertenece el afiliado.

left_join(
  Marcacion_Cohortes, by = "IDEN_AFIL") %>%

```

Se añade informacion del maestro de afiliados.

left_join(

Eafil %>%

dplyr::select(AFIC_DOCUMENTO, AFIC_TIPO_DOCUMENTO, AFIC_SEXO, AFIC_ZONA,
AFIC_REGIMEN, AFIN_NIVEL_SISBEN,

DEPTO_DPN = ID_DTO_DPN, NOM_DEPTO_DPN = NOM_DTO_DPN,

MUN_DPN = ID_M_DPN, NOM_MUN_DPN = NOM_M_DPN,

DEPTO_ATEN = ID_DTO_ATE,

NOM_DEPTO_ATEN = NOM_DTO_ATE,

MUN_ATEN = ID_MUN_ATE,

NOM_MUN_ATEN = NOM_MUN_ATE,

AFIC_ESTADO_AFILIADO, AFIC_ESTADO_BDUA, AFIC_ESTADO_MERCADEO,
AFIF_MODIFICADO, AFIF_RETIRO,

AFIN_NIVEL_CONTRIBUTIVO, FECHA_CORTE_MAESTRO_AFIL = FECHA_CORTE,
MES_MAESTRO_AFIL = MES, ANHO_MAESTRO_AFIL = ANIO) %>%

mutate(IDEN_AFIL = sql("CONCAT(AFIC_TIPO_DOCUMENTO, ' - ', AFIC_DOCUMENTO)")) %>%

mutate(FECHA_CORTE_MAESTRO_AFIL_MES =
sql("DATE_TRUNC(FECHA_CORTE_MAESTRO_AFIL, MONTH)")),

by = c("IDEN_AFIL"="IDEN_AFIL",

"FPRES_MES"="FECHA_CORTE_MAESTRO_AFIL_MES")) %>%

#Se Agrega La fecha de fallecimiento o retiro, y el primer y ultimo estado del afiliado

```
left_join(  
  Agregado_Afil_Primer_Ult_est, by = "IDEN_AFIL"  
) %>%
```

Se agrega informacion de Hospitalizaciones

```
left_join(DetalleHops,  
  by = c("IDEN_AFIL"="IDEN_AFIL",  
        "IDPRESTADOR" = "FCDV_PROVEEDOR",  
        "IDAUTORIZACION"= "FCDN_AUTORIZACION")) %>%  
mutate(FCDV_TOTAL = as.numeric(FCDV_TOTAL)) %>%
```

Se indexan los valores mediante inflacion

```
mutate(VLRSINIESTRO2_Indexado = case_when(  
  ANHO_PRES == 2017 ~ VLRSINIESTRO2*1.353393,  
  ANHO_PRES == 2018 ~ VLRSINIESTRO2*1.300214,  
  ANHO_PRES == 2019 ~ VLRSINIESTRO2*1.260142,  
  ANHO_PRES == 2020 ~ VLRSINIESTRO2*1.214009,  
  ANHO_PRES == 2021 ~ VLRSINIESTRO2*1.194773,  
  ANHO_PRES == 2022 ~ VLRSINIESTRO2*1.131200)) %>%
```

```
mutate(FCDV_TOTAL_Indexado = case_when(  
  ANHO_PRES == 2017 ~ FCDV_TOTAL*1.353393,
```

```
ANHO_PRES == 2018 ~ FCDV_TOTAL*1.300214,  
ANHO_PRES == 2019 ~ FCDV_TOTAL*1.260142,  
ANHO_PRES == 2020 ~ FCDV_TOTAL*1.214009,  
ANHO_PRES == 2021 ~ FCDV_TOTAL*1.194773,  
ANHO_PRES == 2022 ~ FCDV_TOTAL*1.131200)) %>%
```

```
# Se Calcula la Edad a corte de 31 de diciembre de 2022
```

```
mutate(F_CORTE = as.Date('2022-12-31')) %>%
```

```
mutate(Edad = sql("DATE_DIFF(F_CORTE, Fecha_Nacimiento, YEAR)")) %>%
```

```
#Se Agrega el Grupo Etario
```

```
mutate(Grupo_Etario = case_when(  
  between(x = Edad, left = 0, right = 5) ~ "0 - 5",  
  between(x = Edad, left = 5, right = 10) ~ "5 - 10",  
  between(x = Edad, left = 10, right = 15) ~ "10 - 15",  
  between(x = Edad, left = 15, right = 20) ~ "15 - 20",  
  between(x = Edad, left = 20, right = 25) ~ "20 - 25",  
  between(x = Edad, left = 25, right = 30) ~ "25 - 30",  
  between(x = Edad, left = 30, right = 35) ~ "30 - 35",  
  between(x = Edad, left = 35, right = 40) ~ "35 - 40",  
  between(x = Edad, left = 40, right = 45) ~ "40 - 45",  
  between(x = Edad, left = 45, right = 50) ~ "45 - 50",  
  between(x = Edad, left = 50, right = 55) ~ "50 - 55",  
  between(x = Edad, left = 55, right = 60) ~ "55 - 60",
```



```

between(x = Edad, left = 60, right = 65) ~ "60 - 65",
between(x = Edad, left = 65, right = 70) ~ "65 - 70",
between(x = Edad, left = 70, right = 75) ~ "70 - 75",
between(x = Edad, left = 75, right = 80) ~ "75 - 80",
between(x = Edad, left = 80, right = 85) ~ "80 - 85",
Edad > 85 ~ "85 +",

TRUE ~ "SIN CLASIFICAR"

)) %>%

mutate(Grupo_Etario_Class = case_when(
  between(x = Edad, left = 0, right = 5) ~ 1,
  between(x = Edad, left = 5, right = 10) ~ 2,
  between(x = Edad, left = 10, right = 15) ~ 3,
  between(x = Edad, left = 15, right = 20) ~ 4,
  between(x = Edad, left = 20, right = 25) ~ 5,
  between(x = Edad, left = 25, right = 30) ~ 6,
  between(x = Edad, left = 30, right = 35) ~ 7,
  between(x = Edad, left = 35, right = 40) ~ 8,
  between(x = Edad, left = 40, right = 45) ~ 9,
  between(x = Edad, left = 45, right = 50) ~ 10,
  between(x = Edad, left = 50, right = 55) ~ 11,
  between(x = Edad, left = 55, right = 60) ~ 12,
  between(x = Edad, left = 60, right = 65) ~ 13,
  between(x = Edad, left = 65, right = 70) ~ 14,
  between(x = Edad, left = 70, right = 75) ~ 15,

```

```
between(x = Edad, left = 75, right = 80) ~ 16,  
between(x = Edad, left = 80, right = 85) ~ 17,  
Edad > 85 ~ 18,  
TRUE ~ 99))
```

```
#SiniestrosCovidProcesada %>% count()
```

```
# Se Crea Vector de grupos etarios para la caracterizacion de la poblacion.
```

```
GE <- SiniestrosCovidProcesada %>%
```

```
distinct(Grupo_Etario, Grupo_Etario_Class) %>%
```

```
filter(Grupo_Etario != "SIN CLASIFICAR") %>%
```

```
arrange(Grupo_Etario_Class) %>%
```

```
dplyr::select(Grupo_Etario) %>%
```

```
collect() %>% pull()
```

```
## 2.3 Preparacion Maestro Afiliados Positivos Covid -----
```

```

Maestro_Afiliado_Siniestrado <- SiniestrosCovidProcesada %>%
  distinct(IDEN_AFIL) %>%
  left_join(

  PositivosCovid %>%

  # Se seleccionan las variables relevantes

  dplyr::select(IDEN_AFIL, Fecha_Entrega_Resultado_min, Fecha_Entrega_Resultado_max,
fallecido) %>%

  # Se deja un unico registro por persona

  mutate(fila_d = sql('ROW_NUMBER() OVER (PARTITION BY IDEN_AFIL)')) %>% filter(fila_d==1)
%>%

  dplyr::select(-c(fila_d)) %>%
  ungroup() %>%

  # Se Cuenta el numero de pruebas positivas por afiliado.

  left_join(

  PositivosCovid %>%

  group_by(IDEN_AFIL) %>%

  summarise(Numero_Pruebas_Positivas = n_distinct(id_toma_muestra_zscovid)) %>%

  ungroup(), by = "IDEN_AFIL"

  ) %>%

  # Se agrega una variable indicadora para marcar a los usuarios con diagnostico positivo

```

```

mutate(Afiliado_Covid = 1), by = "IDEN_AFIL") %>%

# Se reemplazan los "NA" de la variable Afiliado_Covid por ceros.
mutate(Afiliado_Covid = if_else(is.na(Afiliado_Covid), 0, Afiliado_Covid)) %>%

# Se seleccionan solamente aquellos afiliados que tienen un resultado positivo de covid.
filter(Afiliado_Covid == 1) %>%

# Se genera una nueva fecha la cual corresponde al periodo en el que usualmente el paciente se
ha recuperado del covid.
mutate(Fecha_Entrega_Resultado_max_Mas6semanas =
sql("DATE_ADD(Fecha_Entrega_Resultado_max, INTERVAL 6 WEEK)")) %>%

# Se Busca la informacion relacionada a cohortes de a alto costo (COHORTE, FECHA DE
DIAGNOSTICO)
left_join(
  Esin %>%

mutate(IDEN_AFIL = sql("CONCAT(TIPO_DOCUMENTO, ' - ', DOCUMENTO)")) %>%

# Se seleccionan los campos relevantes
dplyr::select(IDEN_AFIL, SINFECHA_SINIESTRO, CONC_NOMBRE) %>%

# Se deja un unico registro por persona (Una sola Cohorte)

```

```
mutate(fila_d = sql('ROW_NUMBER() OVER (PARTITION BY IDEN_AFIL)')) %>% filter(fila_d==1)
%>% dplyr::select(-c(fila_d)) %>% ungroup() %>% #Distinct con .keep_all = T
```

```
mutate(Esin = 1) %>%
```

```
mutate(SINF_FECHA_SINIESTRO = sql("PARSE_DATE('%d/%m/%Y', SINFECHA_SINIESTRO)")),
```

```
by = "IDEN_AFIL") %>%
```

```
# Se reemplazan los "NA" de la variable Esin por ceros.
```

```
mutate(Esin = if_else(is.na(Esin), 0, Esin)) %>%
```

```
mutate(Esin_Postcovid = if_else(SINF_FECHA_SINIESTRO > Fecha_Entrega_Resultado_min, 1, 0))
%>%
```

```
# Se añade el numero de cohortes a las que pertenece el afiliado.
```

```
left_join(
```

```
Marcacion_Cohortes, by = "IDEN_AFIL") %>%
```

```
# Se añade informacion del maestro de afiliados.
```

```
left_join(
```

```
Eafil %>%
```

```
dplyr::select(AFIC_DOCUMENTO, AFIC_TIPO_DOCUMENTO, AFIC_SEXO, AFIC_ZONA,
AFIC_REGIMEN, AFIN_NIVEL_SISBEN,
```

```
DEPTO_DPN = ID_DTO_DPN, NOM_DEPTO_DPN = NOM_DTO_DPN,
```

```
DEPTO_ATEN = ID_DTO_ATE,
```

```

    NOM_DEPTO_ATEN = NOM_DTO_ATE,

    MUN_ATEN = ID_MUN_ATE,

    NOM_MUN_ATEN = NOM_MUN_ATE,

    AFIC_ESTADO_AFILIADO, AFIC_ESTADO_BDUA, AFIC_ESTADO_MERCADEO,
    AFIF_MODIFICADO, AFIF_RETIRO,

    AFIN_NIVEL_CONTRIBUTIVO, FECHA_CORTE_MAESTRO_AFIL = FECHA_CORTE,
    MES_MAESTRO_AFIL = MES, ANHO_MAESTRO_AFIL = ANIO) %>%

    mutate(IDEN_AFIL = sql("CONCAT(AFIC_TIPO_DOCUMENTO, ' - ', AFIC_DOCUMENTO)")) %>%

    mutate(FECHA_CORTE_MAESTRO_AFIL_MES =
    sql("DATE_TRUNC(FECHA_CORTE_MAESTRO_AFIL, MONTH)")),

    by = c("IDEN_AFIL"="IDEN_AFIL")) %>%

```

#Se Agrega La fecha de fallecimiento o retiro, y el primer y ultimo estado del afiliado

```

left_join(

    Agregado_Afil_Primer_Ult_est, by = "IDEN_AFIL"

) %>%

```

Se agrega el numero de cohortes Pre covid.

```

left_join(

    Esin %>%

```

```

    mutate(IDEN_AFIL = sql("CONCAT(TIPO_DOCUMENTO, ' - ', DOCUMENTO)")) %>%

```

Se seleccionan los campos relevantes

```

dplyr::select(IDEN_AFIL, SINFECHA_SINIESTRO, CONC_NOMBRE) %>%

mutate(SINF_FECHA_SINIESTRO = sql("PARSE_DATE('%d/%m/%Y', SINFECHA_SINIESTRO)"))
%>%

left_join(
  PositivosCovid %>%
  distinct(IDEN_AFIL, Fecha_Entrega_Resultado_min),
  by = "IDEN_AFIL"
) %>%

group_by(IDEN_AFIL) %>%

summarise(Numero_CohortesPrecovid =
n_distinct(CONC_NOMBRE[Fecha_Entrega_Resultado_min < SINFECHA_SINIESTRO]),
  Numero_CohortesPostcovid = n_distinct(CONC_NOMBRE[Fecha_Entrega_Resultado_min
> SINFECHA_SINIESTRO])) %>%

ungroup(),

by = "IDEN_AFIL") %>%

mutate(F_CORTE = as.Date('2022-12-31')) %>%

mutate(Edad = sql("DATE_DIFF(F_CORTE, Fecha_Nacimiento, YEAR)")) %>%

#Se Agrega el Grupo Etario

mutate(Grupo_Etario = case_when(
  between(x = Edad, left = 0, right = 5) ~ "0 - 5",
  between(x = Edad, left = 5, right = 10) ~ "5 - 10",

```

```
between(x = Edad, left = 10, right = 15) ~ "10 - 15",
between(x = Edad, left = 15, right = 20) ~ "15 - 20",
between(x = Edad, left = 20, right = 25) ~ "20 - 25",
between(x = Edad, left = 25, right = 30) ~ "25 - 30",
between(x = Edad, left = 30, right = 35) ~ "30 - 35",
between(x = Edad, left = 35, right = 40) ~ "35 - 40",
between(x = Edad, left = 40, right = 45) ~ "40 - 45",
between(x = Edad, left = 45, right = 50) ~ "45 - 50",
between(x = Edad, left = 50, right = 55) ~ "50 - 55",
between(x = Edad, left = 55, right = 60) ~ "55 - 60",
between(x = Edad, left = 60, right = 65) ~ "60 - 65",
between(x = Edad, left = 65, right = 70) ~ "65 - 70",
between(x = Edad, left = 70, right = 75) ~ "70 - 75",
between(x = Edad, left = 75, right = 80) ~ "75 - 80",
between(x = Edad, left = 80, right = 85) ~ "80 - 85",
Edad > 85 ~ "85 +",
```

```
TRUE ~ "SIN CLASIFICAR"
```

```
)) %>%
```

```
mutate(Grupo_Etario_Class = case_when(
  between(x = Edad, left = 0, right = 5) ~ 1,
  between(x = Edad, left = 5, right = 10) ~ 2,
  between(x = Edad, left = 10, right = 15) ~ 3,
  between(x = Edad, left = 15, right = 20) ~ 4,
  between(x = Edad, left = 20, right = 25) ~ 5,
```



```

between(x = Edad, left = 25, right = 30) ~ 6,
between(x = Edad, left = 30, right = 35) ~ 7,
between(x = Edad, left = 35, right = 40) ~ 8,
between(x = Edad, left = 40, right = 45) ~ 9,
between(x = Edad, left = 45, right = 50) ~ 10,
between(x = Edad, left = 50, right = 55) ~ 11,
between(x = Edad, left = 55, right = 60) ~ 12,
between(x = Edad, left = 60, right = 65) ~ 13,
between(x = Edad, left = 65, right = 70) ~ 14,
between(x = Edad, left = 70, right = 75) ~ 15,
between(x = Edad, left = 75, right = 80) ~ 16,
between(x = Edad, left = 80, right = 85) ~ 17,
Edad > 85 ~ 18,
TRUE ~ 99)) %>%

group_by(IDEN_AFIL) %>%

mutate(Max_maestro = if_else(FECHA_CORTE_MAESTRO_AFIL ==
max(FECHA_CORTE_MAESTRO_AFIL), 1, 0)) %>%

ungroup() %>%

filter(Max_maestro == 1)

source("Script/PreparacionAnalisisDescriptivo.R", encoding = "UTF-8")

```

1. Estadísticas Costo -----

Se compara el costo promedio de atención de un usuario antes y después del covid

para ello se toman los años 2022 y 2019, este último se utiliza como referencia dado que

en 2020 la siniestralidad y los costos disminuyeron debido a la cuarentena.

```
Costo_Atencion_PreVsPostCovid <- SiniestrosCovidProcesada %>%
```

```
  filter(!is.na(Des_Tipo_Siniestro)) %>%
```

```
  # Se seleccionan los años de interés
```

```
  filter(ANHO_PRES %in% c(2019, 2022)) %>%
```

```
  group_by(ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
```

```
  summarise(VLRSINIESTRO2_Indexado = sum(VLRSINIESTRO2_Indexado),
```

```
            Afiliados_Unicos = n_distinct(IDEN_AFIL)) %>%
```

```
  ungroup() %>%
```

```
  collect() %>%
```

```
  mutate(Costo_Medio = VLRSINIESTRO2_Indexado/Afiliados_Unicos) %>%
```

```
  # Se excluyen los siniestros prestados "Pre" y "Durante" el covid del paciente en 2022
```

```
  filter(!(Des_Tipo_Siniestro %in% c("PRE COVID", "DURANTE COVID") & ANHO_PRES == 2022))
```

```
  %>%
```

```
  # Se seleccionan columnas de interés
```

```
  dplyr::select(ANHO_PRES, MES_PRES, Costo_Medio) %>%
```

```
  # Se organiza la data
```

```

arrange(ANHO_PRES) %>%
  pivot_wider(names_from = ANHO_PRES, values_from = Costo_Medio, names_prefix = "Year_")
%>%
  arrange(MES_PRES) %>%
  mutate(Incremento = (Year_2022/Year_2019)-1) %>%
  mutate(NOM_MES = toupper(month(MES_PRES, label = TRUE, abbr = FALSE))) %>%
  dplyr::select(MES_PRES, NOM_MES, 2:4)

```

```
Costo_Atencion_PreVsPostCovid %>%
```

```
  CopyClipboard()
```

```
# Se compara el costo promedio de atencion de un usuario infectado por covid-19
```

```
# para ello se toman los años 2021 y 2022.
```

```
Costo_AtencionCovid_2021Vs2022 <- SiniestrosCovidProcesada %>%
```

```
# Se seleccionan los años de interes
```

```
filter(ANHO_PRES %in% c(2021, 2022)) %>%
```

```
group_by(ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
```

```
summarise(VLRSINIESTRO2_Indexado = sum(VLRSINIESTRO2_Indexado),
```

```
  Afiliados_Unicos = n_distinct(IDEN_AFIL)) %>%
```

```
ungroup() %>%
```

```

collect() %>%

mutate(Costo_Medio = VLR$INIESTRO2_Indexado/Afiliados_Unicos) %>%

# Se Seleccionan los siniestros prestados "Durante" la infeccion por covid del paciente

filter(Des_Tipo_Siniestro == "DURANTE COVID") %>%

# Se seleccionan columnas de interes

dplyr::select(ANHO_PRES, MES_PRES, Costo_Medio) %>%

# Se organiza la data

arrange(ANHO_PRES) %>%

pivot_wider(names_from = ANHO_PRES, values_from = Costo_Medio, names_prefix = "Year_")
%>%

arrange(MES_PRES) %>%

mutate(Incremento = (Year_2022/Year_2021)-1) %>%

mutate(NOM_MES = toupper(month(MES_PRES, label = TRUE, abbr = FALSE))) %>%

dplyr::select(MES_PRES, NOM_MES, 2:4)

Costo_AtencionCovid_2021Vs2022 %>%

write.table("clipboard", col.names = TRUE, row.names=FALSE, sep="\t", dec = ",")

# Se evalua el numero de positivos covid que desarrollaron enfermedades de alto costo posterior
al diagnostico de covid

Valores_AltoCosto <- SiniestrosCovidProcesada %>%

filter(Afiliado_Covid == 1) %>%

```

```

filter(Esin_Postcovid == 1) %>%
filter(FPRES > SINFECHA_SINIESTRO) %>%
group_by(ANHO_PRES) %>%
summarise(VLRSINIESTRO2_Indexado = sum(VLRSINIESTRO2_Indexado),
          Numero_Afiliados = n_distinct(IDEN_AFIL)) %>%
ungroup() %>%
arrange(ANHO_PRES) %>%
collect() %>%
mutate(Total_Afiliados_Positivos = NumPositivos)

```

```
DistrRecuperados <- Maestro_Afiliado_Siniestrado %>%
```

```

filter(fallecido == "NO") %>%

mutate(NOM_DEPTO_ATEN = if_else(NOM_DEPTO_ATEN == "BOGOTA D.C.", "CUNDINAMARCA",
NOM_DEPTO_ATEN)) %>%

group_by(NOM_DEPTO_ATEN) %>%

summarise(Numero_Afiliados=n_distinct(IDEN_AFIL)) %>%

ungroup() %>%

collect() %>%

mutate(Frecuencia_Relativa = Numero_Afiliados/sum(Numero_Afiliados)) %>%

arrange(desc(Frecuencia_Relativa)) %>%

mutate(Frecuencia_Acumulada = cumsum(Frecuencia_Relativa))

```

```
DistrRecuperados %>%
```

```
CopyClipboard(col.names = FALSE)
```

```

DistrRecuperados_Regimen <- Maestro_Afiliado_Siniestrado %>%

  filter(fallecido == "NO") %>%

  mutate(NOM_DEPTO_ATEN = if_else(NOM_DEPTO_ATEN == "BOGOTA D.C.", "CUNDINAMARCA",
  NOM_DEPTO_ATEN)) %>%

  group_by(NOM_DEPTO_ATEN, AFIC_REGIMEN) %>%

  summarise(Numero_Afiliados = n_distinct(IDEN_AFIL)) %>%

  ungroup() %>%

  collect() %>%

  mutate(AFIC_REGIMEN = case_when(

    AFIC_REGIMEN == "S" ~ "SUBSIDIADO",

    AFIC_REGIMEN == "C" ~ "CONTRIBUTIVO",

    TRUE ~ NA_character_

  ))

```

7. Piramides Poblacionales por departamento -----

```

DistrRecuperados_Sexo1 <- Maestro_Afiliado_Siniestrado %>%

  filter(fallecido == "NO") %>%

  mutate(NOM_DEPTO_ATEN = if_else(NOM_DEPTO_ATEN == "BOGOTA D.C.", "CUNDINAMARCA",
  NOM_DEPTO_ATEN)) %>%

  filter(NOM_DEPTO_ATEN %in% c("ATLANTICO", "BOLIVAR", "CORDOBA")) %>%

```

```

group_by(NOM_DEPTO_ATEN, AFIC_SEXO, Grupo_Etario) %>%
summarise(Numero_Afiliados=n_distinct(IDEN_AFIL)) %>%
ungroup() %>%
collect() %>%
filter(Grupo_Etario != "SIN CLASIFICAR") %>%
mutate(Grupo_Etario = factor(Grupo_Etario, levels = GE)) %>%
mutate(AFIC_SEXO = factor(AFIC_SEXO, levels = c("M","F")))

```

```

PlotPiramide_2(Agregado_Caracterizacion = DistrRecuperados_Sexo1,
               GruposEtarios = "Grupo_Etario", Genero = "AFIC_SEXO", NumPersonas =
"Numero_Afiliados",
               Segmentos = "NOM_DEPTO_ATEN", Titulo = "Piramides Poblacionales - Recuperados de
Covid-19", nrow = 1)

```

```

DistrRecuperados_Sexo2 <- Maestro_Afiliado_Siniestrado %>%
filter(fallecido == "NO") %>%
mutate(NOM_DEPTO_ATEN = if_else(NOM_DEPTO_ATEN == "BOGOTA D.C.", "CUNDINAMARCA",
NOM_DEPTO_ATEN)) %>%
filter(NOM_DEPTO_ATEN %in% c("MAGDALENA", "SUCRE", "CESAR")) %>%
group_by(NOM_DEPTO_ATEN, AFIC_SEXO, Grupo_Etario) %>%
summarise(Numero_Afiliados=n_distinct(IDEN_AFIL)) %>%
ungroup() %>%
collect() %>%
filter(Grupo_Etario != "SIN CLASIFICAR") %>%
mutate(Grupo_Etario = factor(Grupo_Etario, levels = GE)) %>%

```

```
mutate(AFIC_SEXO = factor(AFIC_SEXO, levels = c("M","F")))
```

```
PlotPiramide_2(Agregado_Caracterizacion = DistrRecuperados_Sexo2,
```

```
  GruposEtarios = "Grupo_Etario", Genero = "AFIC_SEXO", NumPersonas =  
  "Numero_Afiliados",
```

```
  Segmentos = "NOM_DEPTO_ATEN", Titulo = "Piramides Poblacionales - Recuperados de  
  Covid-19", nrow = 1)
```

```
# 8. Estadísticas Costo Regional -----
```

```
# Se inicializan las listas que contendrán los dataframes de cada regional
```

```
Costo_Atencion_PreVsPostCovid_list <- list()
```

```
Costo_AtencionCovid_2021Vs2022_list <- list()
```

```
Deptos <- c("ATLANTICO", "BOLIVAR", "CORDOBA", "MAGDALENA", "SUCRE", "CESAR")
```

```
## 8.1 Costo PreCovid Vs PostCovid -----
```

```
# Se compara el costo promedio de atención de un usuario antes y después del covid
```

```
# para ello se toman los años 2022 y 2019, este último se utiliza como referencia dado que
```

```
# en 2020 la siniestralidad y los costos disminuyeron debido a la cuarentena.
```

```
for(i in 1:length(Deptos)){
```

```
  Depto <- Deptos[[i]]
```

```
  Costo_Atencion_PreVsPostCovid_list[[i]] <- SiniestrosCovidProcesada %>%
```



```

mutate(NOM_DEPTO_ATEN_temp = Depto) %>%
filter(nombre_dpto_atencion == NOM_DEPTO_ATEN_temp) %>%
# Se seleccionan los años de interes
filter(ANHO_PRES %in% c(2019, 2022)) %>%
group_by(ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
summarise(VLRSINIESTRO2_Indexado = sum(VLRSINIESTRO2_Indexado),
          Afiliados_Unicos = n_distinct(IDEN_AFIL)) %>%
ungroup() %>%
collect() %>%
mutate(Costo_Medio = VLRSINIESTRO2_Indexado/Afiliados_Unicos) %>%

# Se excluyen los siniestros prestados "Pre" y "Durante" el covid del paciente en 2022
filter(!(Des_Tipo_Siniestro %in% c("PRE COVID", "DURANTE COVID") & ANHO_PRES == 2022))
%>%

# Se seleccionan columnas de interes
dplyr::select(ANHO_PRES, MES_PRES, Costo_Medio) %>%

# Se organiza la data
arrange(ANHO_PRES) %>%

pivot_wider(names_from = ANHO_PRES, values_from = Costo_Medio, names_prefix = "Year_")
%>%

arrange(MES_PRES) %>%

mutate(Incremento = (Year_2022/Year_2019)-1) %>%

mutate(NOM_MES = toupper(month(MES_PRES, label = TRUE, abbr = FALSE))) %>%

```

```

dplyr::select(MES_PRES, NOM_MES, 2:4) %>%
mutate(NOM_DEPTO_ATEN = Depto)
}

Costo_Atencion_PreVsPostCovid_depto <- bind_rows(Costo_Atencion_PreVsPostCovid_list)

```

8.2 Costo atencion InterCovid 2021 vs 2022 -----

```

# Se compara el costo promedio de atencion de un usuario infectado por covid-19
# para ello se toman los años 2021 y 2022.

```

```

for(i in 1:length(Deptos)){
  Depto <- Deptos[[i]]

  Costo_AtencionCovid_2021Vs2022_list[[i]] <- SiniestrosCovidProcesada %>%
    mutate(NOM_DEPTO_ATEN_temp = Depto) %>%
    filter(nombre_dpto_atencion == NOM_DEPTO_ATEN_temp) %>%
    # Se seleccionan los años de interes
    filter(ANHO_PRES %in% c(2021, 2022)) %>%

```

```

group_by(ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
summarise(VLRSINIESTRO2_Indexado = sum(VLRSINIESTRO2_Indexado),
          Afiliados_Unicos = n_distinct(IDEN_AFIL)) %>%
ungroup() %>%
collect() %>%
mutate(Costo_Medio = VLRSINIESTRO2_Indexado/Afiliados_Unicos) %>%

# Se Seleccionan los siniestros prestados "Durante" la infeccion por covid del paciente
filter(Des_Tipo_Siniestro == "DURANTE COVID") %>%
# Se seleccionan columnas de interes
dplyr::select(ANHO_PRES, MES_PRES, Costo_Medio) %>%

# Se organiza la data
arrange(ANHO_PRES) %>%
pivot_wider(names_from = ANHO_PRES, values_from = Costo_Medio, names_prefix = "Year_")
%>%
arrange(MES_PRES) %>%
mutate(Incremento = (Year_2022/Year_2021)-1) %>%
mutate(NOM_MES = toupper(month(MES_PRES, label = TRUE, abbr = FALSE))) %>%
dplyr::select(MES_PRES, NOM_MES, 2:4) %>%
mutate(Depto)
}

Costo_AtencionCovid_2021Vs2022_depto <- bind_rows(Costo_AtencionCovid_2021Vs2022_list)

```

```

Utilizaciones_Globales_PreVsPost <- SiniestrosCovidProcesada %>%
  filter(ANHO_PRES %in% c(2019, 2022)) %>%
  filter(Des_Tipo_Siniestro %in% c("PRE COVID", "POST COVID")) %>%
  group_by(IDEN_AFIL, ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
  summarise(Utilizaciones = n()) %>%
  ungroup() %>%
  group_by(ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
  summarise(Utilizaciones_Promedio = sql("AVG(Utilizaciones)")) %>%
  ungroup() %>%
  collect() %>%

# Se excluyen los siniestros prestados "Pre" y "Durante" el covid del paciente en 2022
filter(!(Des_Tipo_Siniestro %in% c("PRE COVID", "DURANTE COVID") & ANHO_PRES == 2022))
%>%

# Se seleccionan columnas de interes

dplyr::select(ANHO_PRES, MES_PRES, Utilizaciones_Promedio) %>%
  arrange(ANHO_PRES) %>%
  pivot_wider(names_from = ANHO_PRES, values_from = Utilizaciones_Promedio, names_prefix =
"Year_") %>%
  arrange(MES_PRES) %>%
  mutate(Incremento = (Year_2022/Year_2019)-1) %>%
  mutate(NOM_MES = toupper(month(MES_PRES, label = TRUE, abbr = FALSE))) %>%
  dplyr::select(MES_PRES, NOM_MES, 2:4)

```

```
Utilizaciones_Globales_PreVsPost %>%
```

```
  CopyClipboard()
```

```
Utilizaciones_Globales_Intra <- SiniestrosCovidProcesada %>%
```

```
  filter(ANHO_PRES %in% c(2021, 2022)) %>%
```

```
  group_by(IDEN_AFIL, ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
```

```
  summarise(Utilizaciones = n()) %>%
```

```
  ungroup() %>%
```

```
  group_by(ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
```

```
  summarise(Utilizaciones_Promedio = sql("AVG(Utilizaciones)")) %>%
```

```
  ungroup() %>%
```

```
  collect() %>%
```

```
  filter(Des_Tipo_Siniestro == "DURANTE COVID") %>%
```

```
  # Se seleccionan columnas de interes
```

```
  dplyr::select(ANHO_PRES, MES_PRES, Utilizaciones_Promedio) %>%
```

```
  # Se organiza la data
```

```
  arrange(ANHO_PRES) %>%
```

```
  pivot_wider(names_from = ANHO_PRES, values_from = Utilizaciones_Promedio, names_prefix =  
  "Year_") %>%
```

```
  arrange(MES_PRES) %>%
```

```
  mutate(Incremento = (Year_2022/Year_2021)-1) %>%
```

```
  mutate(NOM_MES = toupper(month(MES_PRES, label = TRUE, abbr = FALSE))) %>%
```

```
  dplyr::select(MES_PRES, NOM_MES, 2:4)
```

```
## 8.2 Costo atencion InterCovid 2021 vs 2022 -----
```

```
# Se compara el costo promedio de atencion de un usuario infectado por covid-19
```

```
# para ello se toman los años 2021 y 2022.
```

```
for(i in 1:length(Deptos)){
```

```
  Depto <- Deptos[[i]]
```

```
  Utilizaciones_AtencionCovid_2021Vs2022_list[[i]] <- SiniestrosCovidProcesada %>%
```

```
    mutate(NOM_DEPTO_ATEN_temp = Depto) %>%
```

```
    filter(nombre_dpto_atencion == NOM_DEPTO_ATEN_temp) %>%
```

```
    # Se seleccionan los años de interes
```

```
    filter(ANHO_PRES %in% c(2021, 2022)) %>%
```

```
    group_by(IDEN_AFIL, ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
```

```

summarise(Utilizaciones = n()) %>%
ungroup() %>%
group_by(ANHO_PRES, MES_PRES, Des_Tipo_Siniestro) %>%
summarise(Utilizaciones_Promedio = sql("AVG(Utilizaciones)")) %>%
ungroup() %>%
collect() %>%
filter(Des_Tipo_Siniestro == "DURANTE COVID") %>%
# Se seleccionan columnas de interes
dplyr::select(ANHO_PRES, MES_PRES, Utilizaciones_Promedio) %>%

# Se organiza la data
arrange(ANHO_PRES) %>%
pivot_wider(names_from = ANHO_PRES, values_from = Utilizaciones_Promedio, names_prefix =
"Year_") %>%
arrange(MES_PRES) %>%
mutate(Incremento = (Year_2022/Year_2021)-1) %>%
mutate(NOM_MES = toupper(month(MES_PRES, label = TRUE, abbr = FALSE))) %>%
dplyr::select(MES_PRES, NOM_MES, 2:4) %>%
mutate(NOM_DEPTO_ATEN = Depto)
}

```

```

Utilizaciones_AtencionCovid_2021Vs2022_depto <-
bind_rows(Utilizaciones_AtencionCovid_2021Vs2022_list)

```

```

packages <- c("tidyverse", "lubridate", "readxl", "lemon", "openxlsx", "dbplyr",

```

```

    "fitdistrplus", "actuar",
    "gamlss", "gamlss.dist", "gamlss.tr", "gridGraphics", "ggpubr")
new.packages <- packages[!(packages %in% installed.packages()), "Package"]]
install.packages(new.packages)
for (i in packages) {do.call("library", list(format(i)))}
options(scipen = 999)

```

U=TRUE

```

AfiliadosPostCovid_BaseEstimacion <- SiniestrosCovidProcesada %>%
  filter(nombre_dpto_atencion %in% Deptos) %>%
  filter(Afiliado_Covid == 1) %>%
  filter(ANHO_PRE == 2022) %>%
  group_by(IDEN_AFIL, nombre_dpto_atencion) %>%
  summarise(Utilizaciones = n(),
            Costo_Medio = sum(VLRSINIESTRO2_Indexado)) %>%
  ungroup() %>%
  collect() %>%
  rename(Regional=nombre_dpto_atencion)

```

```

AfiliadosNoCovid_BaseEstimacion <- SiniestrosCovidProcesada %>%

```



```

filter(!is.na(NOM_DEPTO_ATEN)) %>%
filter(NOM_DEPTO_ATEN %in% Deptos) %>%
filter(Afiliado_Covid == 0) %>%
filter(ANHO_PRES == 2022) %>%
group_by(IDEN_AFIL, NOM_DEPTO_ATEN) %>%
summarise(Utilizaciones = n(),
          Costo_Medio = sum(VLRSINIESTRO2_Indexado)) %>%
ungroup() %>%
collect() %>%
rename(Regional=NOM_DEPTO_ATEN)

```

```
Df_Mod <- AfiliadosPostCovid_BaseEstimacion
```

```
#vector de Regionales
```

```
Regionales_vector <- Df_Mod %>%
```

```
distinct(Regional) %>%
```

```
pull(Regional)
```

```
Regionales_vector
```

```

# Funcion para extraer percentiles y observaciones aleatorias de las distribuciones
SalidaModelos <- function(modelos_ajustados, tipo_salida){

# Funcion para calcular los percentiles y los valores esperados de los modelos
# seleccionados. Devuelve una lista de vectores, uno por cada modelo.
# modelos_ajustados: es una lista con objetos gamlss que contienen los modelos
# ajustados
# tipo_salida: es una cadena de caracteres que especifica la salida para los
# modelos ajustados, es decir, "percentiles", "muestra_aleatoria" o "valores_esperados"

l <- lapply(modelos_ajustados, function(m){
  dist <- m$family[1]
  par_nombres <- m$parameters

  ll <- m$parameters %>% length()
  # par_estimados <- list()
  # for (i in 1:ll){
  #   par_estimados <- append(par_estimados, m[par_nombres[[i]] %>% as.character()])
  # }

  par_estimados <- lapply(par_nombres, function(p){ fitted(m, parameter=p)[1]})
  names(par_estimados) <- par_nombres

  qdist <- get(paste0("q", dist))

```

```

rdist <- get(paste0("r",dist))

if(tipo_salida=="percentiles"){
  q <- do.call(qdist, c(list(p=seq(0.01, .99,length=100)), par_estimados))
  return(q)
}

if(tipo_salida=="muestra_aleatoria"){
  r <- do.call(rdist, c(list(n=100e+06), par_estimados))
  return(r)
}

if(tipo_salida=="valores_esperados"){
  r <- do.call(rdist, c(list(n=10e+06), par_estimados))
  mm <- moments::moment(r, order = 1)
  return(mm)
}

})

names(l) <- sapply(modelos_ajustados, function(m) m$family[1])
return(l)
}

SalidaModelos_2 <- function(modelos_ajustados, tipo_salida){
  # Función para calcular los percentiles y los valores esperados de los modelos

```

```

# seleccionados. Devuelve una lista de vectores, uno por cada modelo.
# modelos_ajustados: es una lista con objetos gamlss que contienen los modelos
# ajustados
# tipo_salida: es una cadena de caracteres que especifica la salida para los
# modelos ajustados, es decir, "percentiles", "muestra_aleatoria" o "valores_esperados"

l <- lapply(modelos_ajustados, function(m){
  dist <- m$family[1]
  par_nombres <- m$parameters

  ll <- m$parameters %>% length()
  # par_estimados <- list()
  # for (i in 1:ll){
  #   par_estimados <- append(par_estimados, m[par_nombres[[i]] %>% as.character()])
  # }
  par_estimados <- lapply(par_nombres, function(p){ fitted(m, parameter=p)[1]})
  names(par_estimados) <- par_nombres

  qdist <- get(paste0("q", dist))
  rdist <- get(paste0("r", dist))

  if(tipo_salida == "percentiles"){
    q <- do.call(qdist, c(list(p=seq(0.01,.99,length=100)), par_estimados))
    return(q)
  }
}

```

```

}
if(tipo_salida=="muestra_aleatoria"){
  r <- do.call(rdist, c(list(n=100e+06), par_estimados))
  return(r)
}
if(tipo_salida=="valores_esperados"){
  r <- do.call(rdist, c(list(n=10e+06), par_estimados))
  mm <- moments::moment(r, order = 1)
  return(mm)
}

})

names(l) <- sapply(modelos_ajustados, function(m) m$family[2])
return(l)
}

# Funcion para calcular el error cuadratico medio
ECM <- function(q_emp, q_teo){
  N <- length(q_emp)
  sum((q_emp-q_teo)^2)
}

# Se generan las distribuciones truncadas

```

```

Discrete_Distr <- c("PO", "GEOM", "NBI")

Truncated_Discrete_Distr <- c() # Vector que contiene los nombres con las funciones truncadas
for (i in Discrete_Distr){
  do.call(gen.trun, list(par = c(0), family = i, name = "_tr", type = "left"))
  Truncated_Discrete_Distr <- append(Truncated_Discrete_Distr, paste0(i, "_tr"))
}

# Lista de DataFrame de observacions por cobertura -----

Lista_df_coberturas <- list()
for(i in Regionales_vector){

  Reg <- i

  # Contiene las series de datos de lasutilizaciones por cobertura en el a?o seleccionado para las
  estimaciones

  observaciones_U <- Df_Mod %>%
    filter(Regional == Reg)

  Lista_df_coberturas[[i]] <- observaciones_U
}

```

```

# Lista de DataFrames de los criterios de informacion de cada cobe -----

# Se Genera la lista de criterios de seleccion y del objeto gamlssML para cada cobertura
# Se definen las distribuciones discretas a utilizar (Ver documentacion del paquete "gamlss.dist")
Discrete_Distr <- c("PO", "GEOM", "NBI")

# Se truncan las distribuciones seleccionadas excluyendo el cero.
# Esta opcion genera "nuevas distrubuciones" las cuales quedan incluidas en el paquete de forma
temporal
# y pueden ser utilizadas como cualquier otra funcion de R

Truncated_Discrete_Distr <- c() # Vector que contiene los nombres con las funciones truncadas
for (i in Discrete_Distr){
  do.call(gen.trun, list(par = c(0), family = i, name = "_tr", type = "left"))
  Truncated_Discrete_Distr <- append(Truncated_Discrete_Distr, paste0(i, "_tr"))
}

```

```

Criterios <- list()

Distribuciones_cob <- list()

for(i in Regionales_vector){

  cob_df <- Lista_df_coberturas[[i]]

  ob<-if (U==TRUE){cob_df$Utilizaciones} else {cob_df$UtilizacionesEquivalentes}

  # Se define define y ajusta un el modelo inicial gamlss
  m_U <- gamlssML(ob, family = PO_tr)

  # Se ajustan las distribuciones adicionales
  r_U <- chooseDist(m_U, type = "extra", extra = Truncated_Discrete_Distr)
  colnames(r_U) <- c("AIC", "Chi square", "BIC")
  modelos_U <- getOrder(r_U, column = 1)

  modelos_ajustados_U <- lapply(names(modelos_U), function(x){
    gamlssML(ob, family =as.name(x))
  })

  Distribuciones_cob[[i]] <- modelos_ajustados_U
}

```



```

names(modelos_ajustados_U) <- names(modelos_U)

r_U <- r_U[names(modelos_U), ]

df_ <- as.data.frame(r_U)
df_ <- rownames_to_column(df_, "Distr")

Criterios[[i]] <- df_

}

# Se guarda la lista de los Q-Q plot de cada cobertura
Lista_qq_plot <- list()
ListaAjusteDistr <- list()
for(i in Regionales_vector){

  observaciones <-if (U==TRUE) {Lista_df_coberturas[[i]]$Utilizaciones} else
  {Lista_df_coberturas[[i]]$U_ajustado}

  modelos<-modelos_U

  modelos_ajustados<- Distribuciones_cob[[i]]

```

```
r1<- Criterios[[i]]
```

```
# Evaluar bondad de ajuste de las funciones de densidad
```

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

```
lapply(names(modelos), function(x){
```

```
  histDist(observaciones, density = TRUE, nbins=7, family=x, line.col="green", col.hist = "red")
```

```
})
```

```
modelos <- names(modelos)
```

```
for (i in 1:length(modelos)) {
```

```
  histDist(observaciones, density = TRUE, nbins = 7, family = names(modelos)[i], line.col = "green",  
col.hist = "red")
```

```
  grid.echo()
```

```
  Obj3 <- grid.grab()
```

```
  ListaAjusteDistr[[i]] <- ggpubr::ggarrange(Obj3)
```

```
}
```

```
# Percentiles empiricos
```

```
q_emp <- quantile(observaciones, probs=seq(0.01,.99,length=100))
```

```

# Percentiles teoricos

q_teo <- SalidaModelos(modelos_ajustados = modelos_ajustados, tipo_salida = "percentiles")

# ECM

ecm <- lapply(q_teo, FUN = ECM, q_emp=q_emp) %>% unlist()

# 1. Criterios de informacion y ECM

resultados_criterios <- r1 %>% cbind(ecm)

puntuacion <- rowSums(apply(resultados_criterios, MARGIN = 2, FUN = rank))

resultados_criterios2 <- resultados_criterios %>%

  cbind(puntuacion)

resultados_criterios2[order(puntuacion), ]

# Seleccion del top 3 de distribuciones con mejor ajuste

top3 <- resultados_criterios2[order(puntuacion), ][1:3, ] %>% row.names()

q_teo_top3 <- q_teo[top3]

modelos_ajustados_top3 <- modelos_ajustados[top3]

```

```
# 3. QQ-plot
```

```
par(mfrow = c(1,3))
```

```
for(j in 1:length(q_teo_top3)){
```

```
  plot(x = q_teo_top3[[j]], y = q_emp, type = 'p', cex = 0.7,  
        main = paste0('Q-Q plot ', top3[j], " Util por Usuario"),  
        xlab = 'Cuantiles teoricos', ylab = 'Cuantiles empiricos',)  
  abline(0 , 1, col = 2, lty = 2, lwd = 2)
```

```
}
```

```
grid.echo()
```

```
Obj2 <- grid.grab()
```

```
Lista_qq_plot[[i]] <- ggpubr::ggarrange(Obj2)
```

```
}
```

```
# Lista de DataFrames de los Parametros de la distribucion selecci -----
```

```

# Se genera una lista de los parametros de las distribuciones seleccionadas

Param_value_list <- list()

param_name_list <- list()

param_df_list <- list()

list_Distr_name <- list()

for(i in Regionales_vector){

  Modelo <- Distribuciones_cob[[i]][[1]]

  param_name_list[[i]] <- Modelo$parameters

  list_Distr_name[[i]] <- Modelo$family[[2]]

  Param_value_list[[i]] <- lapply(Modelo$parameters, function(p){ fitted(Modelo,
parameter=p)[1]})

  param_df_list[[i]] <- tibble(Parametro = param_name_list[[i]],
                             Valor = Param_value_list[[i]] %>% unlist())

}

par_estimados <- lapply(Modelo$parameters, function(p){ fitted(Modelo, parameter=p)[1]})

```

```

# Lista de DataFRames con los valores esperados de cada distribuci -----

# Se genera una lista de dataframes de los valores esperados de la distribuciones seleccionadas

Expected_value_df_List <- list()

Dist_names <- list()

Expected_values <- list()

for(i in Regionales_vector){

  cat(paste0("Calculando Valor Esperado de la Regional: ", i , "\n"))

  Obj <- SalidaModelos_2(modelos_ajustados = Distribuciones_cob[[i]], tipo_salida =
"valores_esperados")

  Dist_names[[i]] <- names(Obj)

  Expected_values[[i]] <- Obj %>% unlist(use.names = FALSE)

  Expected_value_df_List[[i]] <- tibble(Cobertura = i,
                                     Distribucion = Dist_names[[i]],
                                     `Utilizaciones Esperadas` = Expected_values[[i]])
}

```

```
# Vector de valores esperados para la distr seleccionada -----
```

```
# Generar Vector de valores esperados seleccionados
```

```
Expected_values_vector <- list()
```

```
for(i in Regionales_vector){
```

```
  Expected_values_vector[[i]] <- Expected_value_df_List[[i]] %>%
```

```
  #distinct(`Utilizaciones Esperadas`) %>%
```

```
  head(1) %>%
```

```
  pull()
```

```
}
```

```
# Selecciona la distribución con mejor ajuste con base a los criterios.
```

```
Sel_Expected_Value_df <- list()
```

```
for(i in Regionales_vector){
```

```
  Sel_Expected_Value_df[[i]] <- Expected_value_df_List[[i]] %>%
```

```
  #distinct(`Utilizaciones Esperadas`) %>%
```

```
  head(1)
```

```
}
```

```
Sel_Expected_Value_df <- bind_rows(Sel_Expected_Value_df ) %>%
```

```
rename(`Distribución Seleccionada` = Distribucion)
```

```
# Genera un dataset completos con losvalores esperados de cada distribucion
```

```
Expected_value_df <- bind_rows(Expected_value_df_List)
```

```
#Exportar a Word
```

```
rmarkdown::render("Script/PrintResultWord.Rmd")
```