



CARTA DE AUTORIZACIÓN

CÓDIGO

AP-BIB-FO-06

VERSIÓN

1

VIGENCIA

2014

PÁGINA

1 de 2

Neiva, 20 de febrero de 2023

Señores

CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN

UNIVERSIDAD SURCOLOMBIANA

Ciudad

El (Los) suscrito(s):

Osman Andrés Saavedra Bonilla, con C.C. No. 1075314991,

Daniel Felipe Arguello Escobar, con C.C. No. 1075303732,

Autor(es) de la tesis y/o trabajo de grado titulado **Modelamiento de los ingresos a través de un modelo lineal generalizado: Una aplicación al mercado colombiano en el periodo 2019 - 2020**

presentado y aprobado en el año **2023** como requisito para optar al título de **Matemático**;

Autorizo (amos) al CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN de la Universidad Surcolombiana para que, con fines académicos, muestre al país y el exterior la producción intelectual de la Universidad Surcolombiana, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios puedan consultar el contenido de este trabajo de grado en los sitios web que administra la Universidad, en bases de datos, repositorio digital, catálogos y en otros sitios web, redes y sistemas de información nacionales e internacionales "open access" y en las redes de información con las cuales tenga convenio la Institución.
- Permita la consulta, la reproducción y préstamo a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato Cd-Rom o digital desde internet, intranet, etc., y en general para cualquier formato conocido o por conocer, dentro de los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia.
- Continúo conservando los correspondientes derechos sin modificación o restricción alguna; puesto que, de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación del derecho de autor y sus conexos.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, "Los derechos morales sobre el trabajo son propiedad de los autores", los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

Vigilada Mineducación

La versión vigente y controlada de este documento, solo podrá ser consultada a través del sitio web Institucional [www.usco.edu.co](http://www.usco.edu.co), link Sistema Gestión de Calidad. La copia o impresión diferente a la publicada, será considerada como documento no controlado y su uso indebido no es de responsabilidad de la Universidad Surcolombiana.



CARTA DE AUTORIZACIÓN

CÓDIGO

AP-BIB-FO-06

VERSIÓN

1

VIGENCIA

2014

PÁGINA

2 de 2

EL AUTOR/ESTUDIANTE:

Daniel Felipe Arguello Escobar

Firma:

EL AUTOR/ESTUDIANTE:

Osman Andrés Saavedra Bonilla

Firma:



**TÍTULO COMPLETO DEL TRABAJO:** Modelamiento de los ingresos a través de un modelo lineal generalizado: Una aplicación al mercado colombiano en el periodo 2019 - 2020

**AUTOR O AUTORES:**

Primero y Segundo Apellido	Primero y Segundo Nombre
Argüello Escobar	Daniel Felipe
Saavedra Bonilla	Osman Andrés

**DIRECTOR Y CODIRECTOR TESIS:**

Primero y Segundo Apellido	Primero y Segundo Nombre
----------------------------	--------------------------

**ASESOR (ES):**

Primero y Segundo Apellido	Primero y Segundo Nombre
Cangrejo Esquivel	Álvaro Javier

**PARA OPTAR AL TÍTULO DE:** Matemático

**FACULTAD:** Ciencias Exactas y Naturales

**PROGRAMA O POSGRADO:** Matemática aplicada



CIUDAD: Neiva

AÑO DE PRESENTACIÓN: 2023 NÚMERO DE PÁGINAS: 93

TIPO DE ILUSTRACIONES (Marcar con una X):

Diagramas\_\_\_ Fotografías\_\_\_ Grabaciones en discos  Ilustraciones en general\_\_\_ Grabados\_\_\_  
Láminas\_\_\_ Litografías\_\_\_ Mapas\_\_\_ Música impresa\_\_\_ Planos\_\_\_ Retratos\_\_\_ Sin ilustraciones\_\_\_ Tablas  
o Cuadros\_\_\_

SOFTWARE requerido y/o especializado para la lectura del documento: PDF

MATERIAL ANEXO:

PREMIO O DISTINCIÓN (En caso de ser LAUREADAS o Meritoria): Ninguno

PALABRAS CLAVES EN ESPAÑOL E INGLÉS:

Español

Inglés

- |                               |                          |
|-------------------------------|--------------------------|
| 1. Gamma                      | Gamma                    |
| 2. Modelo Lineal Generalizado | Generalized Linear Model |
| 3. Ingresos                   | Income                   |
| 4. Enfoque Clásico            | Classical Approach       |
| 5. Enfoque Bayesiano          | Bayesian Approach        |

RESUMEN DEL CONTENIDO: (Máximo 250 palabras)

En este proyecto, se estudió al grupo de las empresas colombianas que presentan sus estados financieros a la superintendencia financiera comprendidos en el periodo 2019 - 2020, dichas empresas están segmentadas por: microempresa, pequeña empresa, mediana empresa y macroempresa. Se examinó los ingresos obtenidos en esos dos años y se determinó la significancia que tuvieron cada una de ellas frente a la llegada de la pandemia (COVID-19).

La metodología de la investigación es de carácter cuantitativo, es de tipo descriptivo con un diseño no experimental donde se utilizó un modelo lineal generalizado teniendo en cuenta el enfoque clásico y bayesiano para determinar cuál de las dos presentan mejores resultados. Se expusieron 1664 ingresos de empresas colombianas de ambos años tomados de la página web Datos Abiertos Colombia y análogamente 366 datos de la TRM. Los diferentes procedimientos se realizaron utilizando el software estadístico R-studio.

En el análisis al modelo lineal generalizado con respuesta gamma se tiene que las microempresas y



pequeñas empresas presentan un buen ajuste tanto en el enfoque clásico como el bayesiano. Pero para la mediana y macroempresa se observaron pequeñas diferencias en el ajuste de ambos enfoques. Finalmente, se tiene que el enfoque bayesiano genera mejores resultados y además, se observó que los ingresos de las empresas en Colombia reportaron un decaimiento en el año 2020 frente al año 2019. Es decir, que la pandemia si afecto significativamente a las empresas colombiana y especialmente a las microempresas y pequeñas empresas.

**ABSTRACT:** (Máximo 250 palabras)

In this project, we studied the group of Colombian companies that submit their financial statements to the financial superintendence in the period 2019 - 2020, these companies are segmented by: microenterprise, small enterprise, medium enterprise, and microenterprise. The income obtained in those two years was examined and the significance that each of them had in the face of the arrival of the pandemic (covid-19) was determined.

The research methodology is quantitative, descriptive with a non-experimental design where a generalized linear model was used taking into account the classical and Bayesian approach to determine which of the two presented better results. A total of 1664 incomes of Colombian companies from both years, taken from the Colombia open data web page, and similarly 366 data from the TRM, were exposed. The different procedures were carried out using the statistical software R-studio.

In the analysis of the generalized linear model with gamma response, micro and small companies show a good fit in both the classical and Bayesian approaches. However, for the medium and macro enterprises, small differences were observed in the fit of both approaches. Finally, we have that the Bayesian approach generates better results and, in addition, it was observed that the income of companies in Colombia reported a decline in 2020 compared to 2019. In other words, the pandemic did significantly affect Colombian companies, especially micro and small companies.

**APROBACION DE LA TESIS**

Nombre Presidente Jurado:

Firma:

Nombre Jurado: Leonardo Fabio Medina

Firma:

LEONARDO MEDINA



**UNIVERSIDAD SURCOLOMBIANA  
GESTIÓN SERVICIOS BIBLIOTECARIOS**



**DESCRIPCIÓN DE LA TESIS Y/O TRABAJOS DE GRADO**

<b>CÓDIGO</b>	<b>AP-BIB-FO-07</b>	<b>VERSIÓN</b>	<b>1</b>	<b>VIGENCIA</b>	<b>2014</b>	<b>PÁGINA</b>	<b>4 de 4</b>
---------------	---------------------	----------------	----------	-----------------	-------------	---------------	---------------

Nombre Jurado: Reynaldo Losada

Firma:

Vigilada Mineducación

La versión vigente y controlada de este documento, solo podrá ser consultada a través del sitio web Institucional [www.usco.edu.co](http://www.usco.edu.co), link Sistema Gestión de Calidad. La copia o impresión diferente a la publicada, será considerada como documento no controlado y su uso indebido no es de responsabilidad de la Universidad Surcolombiana.



UNIVERSIDAD SURCOLOMBIANA - USCO  
FACULTAD DE CIENCIAS EXACTAS Y NATURALES  
PROGRAMA MATEMATICA APLICADA

Modelamiento de los ingresos a través de un modelo lineal  
generalizado: Una aplicación al mercado colombiano en el  
periodo 2019 - 2020

Estudiantes

Daniel Felipe Argüello Escobar

Osmán Andrés Saavedra Bonilla

Neiva - Huila  
2022

UNIVERSIDAD SURCOLOMBIANA - USCO  
FACULTAD DE CIENCIAS EXACTAS Y NATURALES  
PROGRAMA MATEMATICA APLICADA

Estudiantes

Daniel Felipe Argüello Escobar

Osmán Andrés Saavedra Bonilla

Asesor: Alvaro Javier Cangrejo Esquivel

Modelamiento de los ingresos a través de un modelo lineal  
generalizado: Una aplicación al mercado colombiano en el  
periodo 2019 - 2020

Trabajo de grado presentado al Pro-  
grama de Pregrado en Matemática  
Aplicada – Departamento de ma-  
temáticas y estadística – Universidad  
Surcolombiana, como requisito par-  
cial para la obtención del título de  
Matemático.

Neiva - Huila  
2022



# Resumen

En este proyecto, se estudió al grupo de las empresas colombianas que presentan sus estados financieros a la superintendencia financiera comprendidos en el periodo 2019 – 2020, dichas empresas están segmentadas por: microempresa, pequeña empresa, mediana empresa y macroempresa. Se examinó los ingresos obtenidos en esos dos años y se determinó la significancia que tuvieron cada una de ellas frente a la llegada de la pandemia (covid-19).

La metodología de la investigación es de carácter cuantitativo, es de tipo descriptivo con un diseño no experimental donde se utilizó un modelo lineal generalizado teniendo en cuenta el enfoque clásico y bayesiano para determinar cual de las dos presentan mejores resultados. Se expusieron 1664 ingresos de empresas colombianas de ambos años tomados de la página web datos abierto Colombia y análogamente 366 datos de la TRM. Los diferentes procedimientos se realizaron utilizando el software estadístico R-studio.

En el análisis al modelo lineal generalizado con respuesta gamma se tiene que las microempresas y pequeñas empresas presentan un buen ajuste tanto en el enfoque clásico como el bayesiano. Pero para la mediana y macroempresa se observaron pequeñas diferencias en el ajuste de ambos enfoques. Finalmente, se tiene que el enfoque bayesiano genera mejores resultados y, además, se observó que los ingresos de las empresas en Colombia reportaron un decaimiento en el año 2020 frente al año 2019. Es decir, que la pandemia sí afectó significativamente a las empresas colombianas y especialmente a las microempresas y pequeñas empresas.

**Palabras clave:** Gamma, Modelo Lineal Generalizado, Ingresos, Enfoque Clásico, Enfoque Bayesiano.

# Abstract

In this project, we studied the group of Colombian companies that submit their financial statements to the financial superintendence in the period 2019 - 2020, these companies are segmented by: microenterprise, small enterprise, medium enterprise, and microenterprise. The income obtained in those two years was examined and the significance that each of them had in the face of the arrival of the pandemic (covid-19) was determined.

The research methodology is quantitative, descriptive with a non-experimental design where a generalized linear model was used taking into account the classical and Bayesian approach to determine which of the two presented better results. A total of 1664 incomes of Colombian companies from both years, taken from the Colombia open data web page, and similarly 366 data from the TRM, were exposed. The different procedures were carried out using the statistical software R-studio.

In the analysis of the generalized linear model with gamma response, micro and small companies show a good fit in both the classical and Bayesian approaches. However, for the medium and macro enterprises, small differences were observed in the fit of both approaches. Finally, we have that the Bayesian approach generates better results and, in addition, it was observed that the income of companies in Colombia reported a decline in 2020 compared to 2019. In other words, the pandemic did significantly affect Colombian companies, especially micro and small companies.

**Key words:** Gamma, Generalized Linear Model, Income, Classical Approach, Bayesian Approach.

# Índice general

<b>Resumen</b>	1
<b>Planteamiento del Problema</b>	9
Pregunta de Investigación . . . . .	10
<b>Justificación</b>	11
<b>Objetivos</b>	12
Objetivo general . . . . .	12
Objetivos específicos . . . . .	12
<b>Marco referencial</b>	13
<b>Marco teórico</b>	18
Modelos lineales generalizados . . . . .	18
Los componentes de un modelo lineal generalizado . . . . .	18
La Familia Exponencial . . . . .	20
Estadísticas suficientes y enlaces canónicos . . . . .	25
Principios para reducir los datos . . . . .	29
Principio de suficiencia . . . . .	29
Probabilidad . . . . .	34
Axiomas de la Probabilidad . . . . .	34
Estadística Bayesiana . . . . .	35
Teorema de Bayes . . . . .	35
Distribución a priori y posteriori . . . . .	36
Regiones de confianza . . . . .	37
<b>Metodología</b>	39
Naturaleza de la Investigación . . . . .	39
Fuente de Información . . . . .	39
Población . . . . .	39
Muestra . . . . .	39

<b>Instrumentos</b> . . . . .	40
<b>Procesamiento y análisis de la información</b> . . . . .	40
<b>Enfoque clásico</b> . . . . .	40
<b>Enfoque Bayesiano</b> . . . . .	41
<b>Resultados</b>	<b>43</b>
<b>Conclusiones</b>	<b>70</b>

# Índice de figuras

1. Función de distribución acumulada de las curvas logísticas y de valores extremos. . . . .	26
2. Enlace Box-Cox para algunos valores de $\lambda$ . . . . .	27
3. Enlace Aranda-Ordaz para algunos valores de $\alpha$ . . . . .	28
4. Histograma de los ingresos de las MICROEMPRESAS de Colombia 2019 - 2020 . . . . .	44
5. Comparacion de los histogramas de los ingresos de las MICROEMPRESAS colombiana 2019–2020 . . . . .	45
6. Boxplot de los ingresos de las MICROEMPRESAS colombiana 2019–2020 . . . . .	46
7. Gamma (Enfoque Clásico y Enfoque Bayesiano) VS Densidad de los ingresos de las MICROEMPRESAS colombiana 2019–2020 . . . . .	48
8. Histograma de los ingresos de las PEQUEÑAS EMPRESAS de Colombia 2019 - 2020 . . . . .	50
9. Comparacion de los histogramas de los ingresos de las PEQUEÑAS EMPRESAS colombiana 2019–2020 . . . . .	51
10. Boxplot de los ingresos de las PEQUEÑAS EMPRESAS colombiana 2019–2020 . . . . .	52
11. Gamma (Enfoque Clásico y Enfoque Bayesiano) VS Densidad de los ingresos de las PEQUEÑAS EMPRESAS colombiana 2019–2020 . . . . .	54
12. Histograma de los ingresos de las MEDIANAS EMPRESAS de Colombia 2019 - 2020 . . . . .	57
13. Comparacion de los histogramas de los ingresos de las MEDIANAS EMPRESAS colombiana 2019–2020 . . . . .	58
14. Boxplot de los ingresos de las MEDIANAS EMPRESAS colombiana 2019–2020 . . . . .	59
15. Gamma (Enfoque Clásico y Enfoque Bayesiano) VS Densidad de los ingresos de las MEDIANAS EMPRESAS colombiana 2019–2020 . . . . .	61
16. Histograma de los ingresos de las MACROEMPRESAS de Colombia 2019 - 2020 . . . . .	64
17. Comparacion de los histogramas de los ingresos de las MACROEMPRESAS colombiana 2019–2020 . . . . .	65

18. Boxplot de los ingresos de las MACROEMPRESAS colombiana	
2019–2020 . . . . .	66
19. Gamma (Enfoque Clásico y Enfoque Bayesiano) VS Densidad de	
los ingresos de las MACROEMPRESAS colombiana 2019–2020 . . .	68

# Índice de cuadros

1. Características de algunas distribuciones univariantes comunes en la familia exponencial . . . . .	23
2. Los enlaces canónicos para las distribuciones en la familia exponencial	25
3. Descriptiva de los ingresos de las MICROEMPRESAS de Colombia	43
4. Estimaciones de los parámetros con respecto el año 2019 enfoque clásico . . . . .	46
5. Estimaciones de los parámetros con respecto el año 2020 enfoque clásico . . . . .	47
6. Estimación de los parámetros enfoque bayesiano . . . . .	47
7. Información del modelo para los ingresos de las MICROEMPRESAS año 2019 . . . . .	48
8. Información del modelo para los ingresos de las MICROEMPRESAS año 2020 . . . . .	49
9. Descriptiva de los ingresos de las PEQUEÑAS EMPRESAS de Colombia . . . . .	49
10. Estimaciones de los parámetros con respecto el año 2019 enfoque clásico . . . . .	53
11. Estimaciones de los parámetros con respecto el año 2020 enfoque clásico . . . . .	53
12. Estimación de los parámetros enfoque bayesiano . . . . .	53
13. Información del modelo para los ingresos de las PEQUEÑAS EMPRESAS año 2019 . . . . .	55
14. Información del modelo para los ingresos de las PEQUEÑAS EMPRESAS año 2020 . . . . .	55
15. Descriptiva de los ingresos de las MEDIANAS EMPRESAS de Colombia . . . . .	56
16. Estimaciones de los parámetros con respecto el año 2019 enfoque clásico . . . . .	59
17. Estimaciones de los parámetros con respecto el año 2020 enfoque clásico	60
18. Estimación de los parámetros enfoque bayesiano . . . . .	60

19. Información del modelo para los ingresos de las MEDIANAS EMPRESAS año 2019	62
20. Información del modelo para los ingresos de las MEDIANAS EMPRESAS año 2020	62
21. Descriptiva de los ingresos de las MACROEMPRESAS de Colombia	63
22. Estimaciones de los parámetros con respecto el año 2019 enfoque clásico	66
23. Estimaciones de los parámetros con respecto el año 2020 enfoque clásico	67
24. Estimación de los parámetros enfoque bayesiano	67
25. Información del modelo para los ingresos de las MACROEMPRESAS año 2019	69
26. Información del modelo para los ingresos de las MACROEMPRESAS año 2020	69



# Planteamiento del Problema

Desde la confirmación del COVID-19 generada en Wuhan, China y tras la llegada al continente americano en enero del 2020, la organización mundial de la salud (OMS) declaró un estado de emergencia sanitaria por el COVID-19. Que afectó toda la dimensión de vida social y económica alrededor de todo el mundo. América Latina ha sido especialmente afectada por la pandemia tanto a nivel sanitario, como a nivel socioeconómico. Debido a esto, grandes cantidades de empresas y negocios han cerrado debido a la cuarentena y las restricciones del distanciamiento social. Pero, antes de la pandemia el país ya tenía problemas de desempleo, deudas externas y crisis fiscal. Lo que generó una prolongación más grave de estos problemas a corto y largo plazo.

La pandemia ha generado grandes problemas económicos, de acuerdo con la CEPAL, durante el 2020 cerraron 2,7 millones de empresas equivalentes al 19% de todas las firmas de la región, se perdieron el 8.7 millones de empleos por culpa de la cuarentena y las restricciones al distanciamiento social. Para ACOPI, esta afectación de la economía se presenta con mayor intensidad en las micro, pequeñas y medianas empresas (MiPymes) que son responsables de 17 millones de empleos en el país. De acuerdo con la Organización Internacional del Trabajo (OIT), el mercado laboral ha sido uno de los más afectados en el mundo puesto que se perdieron 345 millones de empleos y 5,5% del Producto Interno Bruto (PIB) equivalente a los ingresos derivados de la destrucción de empleo. En Colombia, 1,5 millones de personas perdieron sus empleos al comparar octubre 2020 con el mismo mes 2019, Según el DANE el año 2020 presentó una tasa de desempleo del 15,9% y para el año 2019 una tasa del 10,5%, es decir que aumentó un 5,4% en el año 2020. Se presentaron tantos cambios en la economía que incluso, el PIB (producto interno bruto) en Colombia presentó una disminución de -6,8%, después de que en el 2019 se registrara un crecimiento del 2,9%.

En Colombia el sector empresarial se encuentra segmentado por microempresas, pequeñas empresas, medianas empresas y macroempresas. La mediana y la macroempresa generan dos tercios de empleos y aportan al producto interno bruto. La micro y pequeña empresa, aunque no generan grandes ingresos anuales son las que

están mas presentes en el país, pues en Colombia el 92 % del sector empresarial son micro y pequeñas empresas y todas juntas generan el 35 % del producto interno bruto de Colombia. Es decir que es fundamental para el desarrollo y crecimiento del país.

Todo el sector empresarial contribuye al 90 % del sector nacional y generan el 80 % de empleos. Durante la pandemia estas empresas fueron afectadas drásticamente, aunque no se tiene unas cifras exactas de números de empresas que cerraron en el año 2019 y 2020. La cámara y comercio de Bogotá revelo que para el año 2020 se liquidaron 37.000 empresas y 96 % pertenecen al sector empresarial. En Colombia las microempresas fueron las más afectadas, según el DANE para el año 2020 se estimó que 509.370 microempresas cerraron a nivel nacional.

Las afectaciones socioeconómicas por la pandemia a nivel nacional requieren de estrategias para recuperar el empleo y los ingresos del sector productivo. En Colombia el gobierno nacional planifico y evaluó alternativas sectorizadas, con la intención de generar condiciones plenas y sostenibles para la recuperación de la economía. Debido a que en Colombia se encuentra una gran mayoría de micro y pequeñas empresas, las acciones de fortalecimiento de post pandemia deber ser enfocada principalmente a estas a través de nuevas estrategias de mercadeo y nuevas condiciones económicas.

En consecuencia, podemos decir que el impacto económico ha afectado gravemente en el sector empresarial de Colombia y su impacto es diferentes para la micro, pequeña, mediana y gran empresa debido a

- Impacto directo de la producción
- Interrupción de la cadena de suministro y del mercado
- Repercusiones financieras en las empresas y los mercados financieros

Por consiguiente, se quiere mirar el comportamiento de los ingresos de las empresas en Colombia durante el periodo 2019 – 2020. Es apropiado plantearse como pregunta de investigación Existieron cambios significativos en los ingresos de las empresas en Colombia a causa de la pandemia (Covid–19)?

## **Pregunta de Investigación**

¿Existieron cambios significativos en los ingresos de la empresas en Colombia a causa de la pandemia (Covid–19)?.

# Justificación

La afectación en el sector empresarial de Colombia por la pandemia del COVID–19 ha sido negativa para todos los departamentos y para la economía del país. Los sectores económicos tuvieron mayor restricción de operación por las obligaciones impuestas del gobierno nacional (cuarentena obligatoria, restricciones al distanciamiento social, entre otros) que produjo una disminución intensiva de los bienes y servicios.

Estas repercusiones negativas se extenderán a corto y largo plazo afectando principalmente al sector empresarial llevando acabo afectaciones en el mercado laboral y al sostenimiento de los hogares colombianos.

La importancia de la investigación se justifica teniendo en cuenta que la pandemia del COVID–19 ha generado un cambio drástico en la economía global y el entorno empresarial. En muchos casos –como en los sectores de servicios, comercio y turismo– los impactos fueron inmediatos y brutales, mientras que otras áreas terminaron teniendo una mayor supervivencia en estos primeros meses y algunas incluso se beneficiaron por la naturaleza de los servicios. (Inova, 2020)

El presente trabajo busca brindar información sobre las afectaciones que contrajeron la empresa por la llegada del COVID–19 como:

- Afectación según el sector empresarial
- Niveles de disminución de ingresos del 2020 frente al 2019
- Expectativa de apoyo a las empresas más afectadas

La recolección y análisis de información se enfoca en la orientación de los ingresos generados por las empresas colombianas en el año 2020 (Durante el COVID–19) y compararlos con los del año 2019 (Antes del COVID–19). Para determinar el impacto social económico que genero la pandemia en cada una de las empresas que conforma el sector empresarial, e incentivar a prevenir las repercusiones de corto y largo plazo presentadas por el COVID–19, que podrían conllevar a que muchas de las empresas presentes decidan cerrar parcial o definitivamente sus operaciones y otros se reinventaron o innovaron para no salir del mercado.

# Objetivos

## Objetivo general

Analizar y modelar mediante un modelo lineal generalizado los ingresos de las empresas en Colombia utilizando estimación clásica y bayesiana.

## Objetivos específicos

1. Estimar los parámetros del modelo a través del enfoque clásico.
2. Estimar los parámetros del modelo a través del enfoque bayesiano
3. Comparar los resultados obtenidos bajo las diferentes aproximaciones.

# Marco referencial

(**Simbasica Numpaqué & Bermúdez Rubio**, 2019) trabajaron en los ingresos mensuales de los bogotanos, aplicando un modelo lineal generalizado con distribución Tweedie, distribución Inversa Gaussiana ajustada a cero y distribución Gamma ajustada a cero dada la ventaja que tiene función de admitir ceros en su variable respuesta de tipo continuo no negativa, pudiendo ser utilizada en el análisis de los ingresos mensuales de los bogotanos realizada en el año 2017 a sus 20 localidades, encuestando a más de 77025 hogares, donde se examina el impacto que tiene las variables explicativas en la variable respuesta ingresos mensuales promedio, del cual proporciona una visión sobre la situación laboral y esto sirva como estrategia del distrito en políticas públicas enfocado a mejorar el bienestar socioeconómico de los bogotanos. Teniendo en cuenta que, de todas las variables explicativas usadas en el modelo, el nivel de educación es el que tiene un mayor peso a la hora de estimar los ingresos (p. 1). (13)

(**Hernández & Lasso**, 2003) trabajaron en los efectos del salario mínimo legal (SML) sobre el empleo, en una función de demanda de trabajo, teniendo en cuenta el efecto sustitución y el efecto ingreso inducidos por el incremento del salario mínimo. Para empleados cubierto por el SML (jóvenes y no calificados) y los no cubiertos (adultos y calificados), para el periodo 1984 – 2000, dentro de un modelo estándar de demanda de factores (p. 131). Una de las principales conclusiones del estudio es que el efecto neto del SML, sobre el empleo es positivo, tanto para la población cubierta como la no cubierta por este, cuando se considera en conjunto el efecto sustitución y el efecto ingresos. Por otro lado, se muestra que la demanda de trabajo está determinada por el ciclo económico más que por cambio en los precios de los factores de producción (p. 117). (6)

(**Parra**, 1999) trabajo en el campo de la distribución del ingreso se ha incrementado en los últimos años, por la necesidad de determinar el impacto de las medidas recientes de política sobre el bienestar de la economía (p. 135). También es la necesidad de hacer énfasis en la diferenciación al analizar los cambios de la distribución de trabajadores calificados y no calificados, puesto que cada uno presenta procesos completamente diferentes. Este trabajo de investigación es analizar la evolución

de la distribución del ingreso de Colombia en los últimos años desde un enfoque alternativo. Para esto se establece unos hechos básicos que revelan su comportamiento durante este periodo, a partir de una metodología no-paramétrica basada en herramientas estadísticas que permiten identificar los cambios relevantes de la distribución (p. 136).<sup>(9)</sup>

(**Sánchez**, 2005) hizo una recopilación de trabajos recientes sobre los temas de distribución del ingreso, pobreza y política social. En la década de los noventa hubo aumento de la concentración del ingreso urbano en Colombia, muy similar a las tendencias de otros países de América Latina y los Estados Unidos. Para explicar este hecho se han tejido diversa hipótesis, entre otras, el aumento de la demanda por trabajadores calificados y de su salario relativo, la apertura económica, el lento incremento de la cobertura universitaria, la disminución de la movilidad social, y el aumento de la inestabilidad económica.<sup>(11)</sup>

(**Ocampo et al.**, 2000) trabajaron en el artículo Mercado laboral y distribución del ingreso en Colombia en los años noventa en el cual analiza el comportamiento y los cambios del mercado laboral en Colombia en la década de los noventa. A partir de la información disponible en las encuestas de hogares, lo cual analiza posteriormente la dinámica laboral, por sectores económicos, nivel educativo, y posición ocupacional. Después de resumir la investigación sobre distribución del ingreso para el caso colombiano, cuyo objetivo es determinar los efectos de los cambios observados en las condiciones del mercado laboral sobre la distribución del ingreso y la pobreza, para los sectores urbanos y rurales (p. 53).<sup>(15)</sup>

(**Dávila Aragón et al.**, 2016) trabajo en Cálculo del valor en riesgo operacional mediante redes bayesianas para una empresa financiera para identificar y cuantificadores de riesgo operacional que dan lugar a los fraudes, la interrupción de la actividad productiva y la responsabilidad legal, lo cual es una amenaza constante en cualquier empresa. Para ello se toma un enfoque bayesiano y el modelo causa-efecto dinámico que incorpora una información inicial a priori, lo cual se puede agregar opiniones de experto o intervenciones de especialistas para finalmente calcular el riesgo operacional.<sup>(1)</sup>

(**Chiquito Tigua et al.**, 2020) trabajaron Red bayesiana para analizar los mecanismos de adaptación de los comerciantes antes los retos de covid-19 en particular los comerciantes ecuatorianos, debido a la crisis sanitaria, humana y económica que se vive en Ecuador.

La llegada del covid-19 condujo a que los gobiernos impusieran acciones como, el aislamiento, la cuarentena y el distanciamiento social con el fin de reducir la curva de contagios, lo cual afectó la actividad económica. Pues esto disminuyó la producción o incluso su interrupción total, debido a que los comerciantes se

sustentan en la oferta y demanda de sus comercializaciones, por lo que se creó un trabajo de análisis de sensibilidad a la evidencia en redes bayesianas. es un análisis útil para verificar la volatilidad de los mercados financieros relativos a los comerciantes ecuatoriano. (14)

(Silva Veloza & Sánchez Hernández, 2018) trabajaron en el ESTUDIO DEL IPC EN EL DEPARTAMENTO DEL TOLIMA Y DEPARTAMENTOS ALEDAÑOS MEDIANTE GAMs sobre el índice de precio del consumidor (IPC) es la variación absoluta del precio de la canasta familiar, la cual está definida por el DANE, construido sobre una variante de tipo laspeyres, para un seguimiento del gasto de consumo en los hogares de un país, la cantidad puede variar a nivel regional y por lo tanto en su acumulado nacional. Esta variación puede ser distinta en cada país.

Es decir, el trabajo trata de comparar las IPC del Tolima y los departamentos cercanos (Cauca, Valle, Huila Quindío, Caldas, Risaralda, Cundinamarca), se toma Un modelo lineal generalizado, específicamente en modelo adictivos generalizado (GAMs) con fin de determinar posibles valores influénciales del Tolima y departamentos circunvecino. (12)

(Lee, et al., 2018) realizaron una investigación titulada: "efectos de la relación entre el precio de las vivienda y los ingresos en la elección de la tenencia en Taiwán: pronóstico del rendimiento del modelo lineal generalizado jerárquico y el modelo de regresión logística binaria tradicional" Este estudio examinó los factores que influyen en las elecciones de tenencia de los hogares en diferentes condados y ciudades de Taiwán. Los datos recopilados en la Encuesta sobre el Estado de la Vivienda de la Agencia de Construcción y Planificación del Ministerio del Interior se analizaron mediante modelos lineales generalizados jerárquicos (HGLM). El estudio designó el sector de los hogares como una unidad en el nivel 1 y los condados y ciudades como una unidad en el nivel 2, con la diferencia entre los condados y ciudades representando el 9% de la variación total en las decisiones de alquiler y compra. Con base en los resultados empíricos, la elección de la tenencia se vio afectada positiva y significativamente por factores de nivel 1 como el género, la edad, el nivel educativo, el área per cápita, el número de habitaciones per cápita, los préstamos privados y los ingresos permanentes. La variable de atributo de nivel 2, la relación entre el precio de la vivienda y los ingresos tuvo un efecto negativo significativo en la elección de la tenencia; una relación más alta entre el precio de la vivienda y los ingresos resultó en una mayor preferencia entre los hogares hacia el arrendamiento en sus decisiones de arrendamiento o compra. Con respecto a la comparación de la capacidad de pronóstico, la tasa de acierto de HGLM (90, 10%) fue mayor que la del modelo de regresión logística binaria (87, 26%). En términos de precisión de pronóstico evaluada usando cuatro medidas de asociación,

HGLM superó al modelo de regresión logística binaria tradicional. Basado en una validación cruzada de diez veces, HGLM también mostró una mejor tasa de aciertos que el modelo de regresión logística binaria tradicional, lo que significa que los resultados de la evaluación fueron sólidos y confiables. (7)

(Cingano, 2014) realizaron una investigación titulada: "Tendencias de la desigualdad de ingresos y su impacto en el crecimiento económico."

1. En la mayoría de los países de la OCDE, la diferencia entre ricos y pobres está en su nivel más alto desde hace 30 años. En la actualidad, el 10 % más rico de la población de la zona de la OCDE gana 9,5 veces los ingresos del 10 % más pobre; en la década de 1980 esta proporción era de 7:1 y ha ido aumentando continuamente desde entonces. Sin embargo, el aumento de la desigualdad global de ingresos no se debe (únicamente) al incremento de la proporción de los ingresos superiores: a menudo, los ingresos inferiores crecen mucho más despacio durante los años de prosperidad y caen durante las recesiones, lo que hace que la pobreza de ingresos relativa (y en algunos países, absoluta) esté en el punto de mira de las políticas. Este documento explora si esta evolución puede tener un impacto en los resultados económicos.
2. A partir de datos armonizados que abarcan los países de la OCDE durante los últimos 30 años, el análisis econométrico sugiere que la desigualdad de ingresos tiene un impacto negativo y estadísticamente significativo en el crecimiento posterior. En concreto, lo que más influye es la diferencia entre los hogares con ingresos bajos y el resto de la población. En cambio, no se encuentran pruebas de que los que tienen ingresos altos se alejen del resto de la población, lo que perjudica el crecimiento. El documento también evalúa la "teoría de la acumulación de capital humano," encontrando pruebas de que el capital humano es un canal a través del cual la desigualdad puede afectar al crecimiento. Un análisis basado en microdatos de la Encuesta de Aptitudes de los Adultos (PIAAC) muestra que el aumento de las disparidades de ingresos deprime el desarrollo de las aptitudes entre los individuos con antecedentes educativos parentales más pobres, tanto en lo que respecta a la cantidad de educación alcanzada (por ejemplo, los años de escolarización), como a su calidad (es decir, el dominio de las aptitudes). Sin embargo, los resultados educativos de los individuos de entornos más ricos no se ven afectados por la desigualdad.
3. De ello se desprende que las políticas de reducción de las desigualdades de renta no sólo deben aplicarse para mejorar los resultados sociales, sino también para sostener el crecimiento a largo plazo. Las políticas de redistribución a través de los impuestos y las transferencias son una herramienta clave para



garantizar que los beneficios del crecimiento se distribuyan más ampliamente, y los resultados sugieren que no es necesario esperar que perjudiquen el crecimiento. Pero también es importante promover la igualdad de oportunidades en el acceso y la calidad de la educación. Esto implica centrarse en las familias con niños y jóvenes

– ya que es cuando se toman las decisiones sobre la acumulación de capital humano

– promover el empleo de los grupos desfavorecidos mediante políticas activas del mercado de trabajo, ayudas para el cuidado de los niños y prestaciones en el trabajo. (3)

(Díaz Prieto et al., 2018) realizaron una investigación titulada: "Estimación de un modelo lineal generalizado mixto para datos de conteo con exceso de ceros." En este documento se presentan los análisis realizados a partir del ejercicio de modelación de la cantidad de incendios forestales registrados en los 212 municipios que componen los departamentos de Antioquia y Santander durante el año 2013. Para dicho análisis se empleó el método de modelos lineales generalizados mixtos inflados de ceros en el que se estimó el efecto de diferentes variables climáticas y geográficas medidas en cada municipio. El proceso de estimación se realizó mediante modelos aditivos generalizados de ubicación, escala y forma, en el cual se obtuvieron los parámetros mediante estimaciones lineales generalizadas, estimaciones no paramétricas y una estimación de retardo espacial mediante modelos SAR. En la cual concluyo lo siguiente:

1. Las variables analizadas tienen efectos significativos sobre la cantidad de incendios registrados en cada municipio. Ya que sólo se evaluó un año, es posible expandir esta investigación, tanto a otros años, como a otras regiones del país.
2. La metodología usada de modelos mixtos, incluyendo estimaciones no paramétricas y de autocorrelación espacial, demostró ser mejor que un modelo lineal generalizado Poisson cero inflado, ya que las estimaciones presentaron mejores resultados en los criterios de comparación.
3. Las variables analizadas son estadísticamente significativas, lo que sugiere que las estancias gubernamentales encargadas de la gestión del riesgo pueden enfocar sus esfuerzos en función de estas variables hacia la prevención de futuros siniestros ambientales. (4)

# Marco teórico

## Modelos lineales generalizados

Estudiaremos una extensión de los modelos lineales a una familia más general, propuesta por (Nelder & Wedderburn, 1997), denominada **Modelos Lineales Generalizados (MLG)**. Esta familia unifica tanto los modelos con variables respuesta numéricas como categóricas lo cual lleva a considerar otras distribuciones tales como la binomial, poisson, gamma, etc. además de la normal. (8)

Dado que el modelo lineal general es un buen punto de partida para el estudio de los modelos lineales generalizados, empezamos con una revisión de los principales aspectos del modelo lineal general. (8)

### Los componentes de un modelo lineal generalizado

Los modelos lineales generalizados son una extensión de los modelos lineales clásicos, por lo que estos últimos constituyen un punto de partida adecuado para la discusión. Se supone que un vector de observaciones  $y$  con  $n$  componentes es una realización de una variable aleatoria  $Y$  cuyos componentes se distribuyen independientemente con medias  $\mu$ . La parte sistemática del modelo es una especificación del vector  $\mu$  en términos de un pequeño número de parámetros desconocidos  $\beta_1, \dots, \beta_p$ . En el caso de los modelos lineales ordinarios, esta especificación toma la forma

$$\mu = \sum_1^p x_j \beta_j, \quad (1)$$

donde los  $\beta_s$  son parámetros cuyos valores suelen ser desconocidos y deben estimarse a partir de los datos. Si dejamos que  $i$  indexe las observaciones, la parte sistemática del modelo puede escribirse

$$E(Y_i) = \mu_i = \sum_1^p x_{ij} \beta_j; \quad i = 1, \dots, n, \quad (2)$$

donde  $x_{ij}$  es el valor de la  $j$ th covariable para la observación  $i$ . En notación matricial (donde  $\mu$  es  $n \times 1$ ,  $X$  es  $n \times p$  y  $\beta$  es  $p \times 1$ ) podemos escribir

$$\mu = X\beta$$

donde  $X$  es la matriz del modelo y  $\beta$  es el vector de parámetros. esto completa la especificación de la parte sistemática del modelo. (8)

Para la parte aleatoria suponemos independencia y varianza constante de los errores. Estas suposiciones son fuertes y es necesario comprobarlas, en la medida de lo posible, a partir de los propios datos. Del mismo modo, la estructura de la parte sistemática supone que conocemos las covariables que influyen en la media y que podemos medirlas eficazmente sin errores; esta suposición también debe comprobarse, en la medida de lo posible. (8)

Una especialización más del modelo implica la suposición más fuerte de que los errores siguen una distribución gaussiana o normal con varianza constante  $\sigma^2$ .

Así, podemos resumir el modelo lineal clásico en la forma:

Los componentes de  $Y$  son variables normales independientes con varianza constante  $\sigma^2$  y

$$E(Y) = \mu \quad \text{donde} \quad \mu = X\beta \quad (3)$$

### La generalización

Para simplificar la transición a los modelos lineales generalizados, reorganizaremos (3) ligeramente para producir la siguiente especificación en tres partes:

1. El componente aleatorio: el componente de  $Y$  tiene distribuciones normales independientes con  $E(Y) = \mu$  y varianza constante  $\sigma^2$ ;
2. El componente sistemático: las covariables  $x_1, x_2, \dots, x_p$  producen un predictor lineal  $\eta$  dado por

$$\eta = \sum_1^p x_j \beta_j;$$

3. El vínculo entre los componentes aleatorio y sistemático:

$$\mu = \eta.$$

Esta generalización introduce un nuevo símbolo  $\eta$  para el predictor lineal y el tercer componente especifica entonces que  $\mu$  y  $\eta$  son de hecho idénticos. Si escribimos

$$\eta_i = g(\mu_i)$$

entonces  $g(\cdot)$  se llamará función de enlace. En esta formulación, los modelos lineales clásicos tienen una distribución Normal (o Gaussiana) en el componente 1 y la función de identidad para el enlace en el componente 3. Los modelos lineales generalizados permiten dos extensiones; en primer lugar, la distribución en el componente 1 puede proceder de una familia exponencial distinta de la Normal y, en segundo lugar, la función de enlace en el componente 3 puede convertirse en cualquier función diferenciable monótona.

En primer lugar, examinamos la hipótesis de distribución ampliada. [\[8\]](#)

## La Familia Exponencial

Como se ha mencionado anteriormente, los modelos lineales generalizados suponen que la variable de respuesta tiene una distribución que pertenece a una familia particular, la familia exponencial. La definición que presentaremos aquí es la adecuada para los modelos de la variable de respuesta que nos interesa considerar en el ámbito de los MLG. Véase, por ejemplo, Cox y Hinkley (1974), para una definición más general de la familia exponencial de  $k$  parámetros y sus propiedades. [\[8\]](#)

**Definición 1** (Familia exponencial). Se dice que una variable aleatoria  $Y$  tiene una distribución que pertenece a la familia de dispersión exponencial (o simplemente familia exponencial) si su función de densidad de probabilidad (f.d.p.) o función de masa de probabilidad (f.m.p.) puede escribirse de la forma

$$f_Y(y; \theta; \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (4)$$

donde  $\theta$  y  $\phi$  son parámetros escalares,  $a(\cdot)$ ,  $b(\cdot)$  y  $c(\cdot)$  son reales conocidos funciones. [\[8\]](#)

En la definición anterior,  $\theta$  es la forma canónica del parámetro de localización y  $\phi$  es un parámetro de dispersión que se supone, en general, conocido. En este caso la distribución descrita en (4) forma parte de la familia exponencial uniparamétrica. Cuando se desconoce el parámetro  $\sigma$ , la distribución puede o no formar parte de la familia exponencial biparamétrica, como se define generalmente (véase, por ejemplo, Cox y Hinkley, 1974). También se admite que la función  $b(\cdot)$  es diferenciable, y que el soporte de la distribución no depende de los parámetros. En este caso demostramos que la familia considerada obedece a las condiciones habituales de regularidad. [\[8\]](#)

### Valor medio y varianza

Sea  $l(\theta; \phi; y) = \log f_y(\theta; \phi; y)$  Definamos la función puntuación

$$S(\theta) = \frac{\partial l(\theta; \phi; Y)}{\partial \theta} \quad (5)$$

Se sabe que para las familias regulares tenemos

$$E(S(\theta)) = 0$$

$$E(S^2(\theta)) = E \left[ \left( \frac{\partial l(\theta; \phi; Y)}{\partial \theta} \right)^2 \right] = -E \left[ \frac{\partial l(\theta; \phi; Y)}{\partial \theta} \right] \quad (6)$$

y por lo tanto como, en el caso de que  $f_Y(y; \theta; \sigma)$  esté dada por (4),

$$l(\theta; \phi; Y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

se obtiene

$$S(\theta) = \frac{Y - b'(\theta)}{a(\phi)} \quad \frac{\partial S(\theta)}{\partial \theta} = -\frac{b''(\theta)}{a(\phi)} \quad (7)$$

Donde  $b'(\theta) = \frac{\partial b(\theta)}{\partial \theta}$  e  $b''(\theta) = \frac{\partial^2 b(\theta)}{\partial \theta^2}$   
 Así, a partir de (6) y (7)

$$E(Y) = \mu = a(\phi)E(S(\theta)) + b'(\theta) = b'(\theta) \quad (8)$$

$$\text{var}(Y) = a^2(\phi)\text{var}(S(\theta)) = a^2(\phi)\frac{b''(\theta)}{a(\phi)} = a(\phi)b''(\theta) \quad (9)$$

Vemos así que la varianza de  $Y$  es el producto de dos funciones; una,  $b''(\theta)$ , que depende sólo del parámetro canónico  $\theta$  (y por tanto del valor medio  $\mu$ ), y que se denomina función de varianza que suele representarse por  $V(\mu)$  y otra,  $a(\phi)$ , que sólo depende del parámetro de dispersión  $\phi$ .

En muchas situaciones de interés, observamos que la función  $a(\phi)$  toma la forma

$$a(\phi) = \frac{\phi}{\omega}$$

donde  $\omega$  es una constante conocida, obteniendo así la varianza de  $Y$  como el producto del parámetro de dispersión por una función de sólo el valor medio.

En este caso la función definida en (4) se escribe en la forma

$$f_Y(y; \theta; \phi; \omega) = \exp \left\{ \frac{\omega(y\theta - b(\theta))}{\phi} + c(y, \phi, \omega) \right\} \quad (10)$$

### Casos particulares

Veamos algunos ejemplos de distribuciones conocidas que pertenecen a la familia en estudio. [\(8\)](#)

#### Normal

Sea  $Y$  una variable aleatoria con distribución normal de media  $\mu$  y varianza  $\sigma^2$ ,  $Y \sim N(\mu, \sigma^2)$ . La función de densidad de  $Y$  se expresa en la forma

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\} = \exp\left[\frac{1}{\sigma^2}\left(\mu y - \frac{\mu^2}{2}\right) - \frac{1}{2}\left\{\log 2\pi\sigma^2 + \frac{y^2}{\sigma^2}\right\}\right]$$

donde  $-\infty < \mu, y < \infty$  y  $\sigma^2 > 0$ . Entonces, para  $\theta = \mu$ ,  $b(\theta) = \frac{\theta^2}{2}$ ,  $\phi = \sigma^{-2}$  y  $c(y, \phi) = \frac{1}{2} \log \frac{\phi}{2\pi} - \frac{\phi y^2}{2}$  obtenemos (4). Vemos fácilmente que la función de varianza viene dada por  $V(\mu) = 1$ . (8)

### Poisson

En el caso de  $Y \sim P(\mu)$ , la función de probabilidad viene dada por

$$e^{-\mu} \frac{\mu^y}{y!} = \exp\{y \log \mu - \mu - \log y!\}$$

donde  $\mu > 0$  e  $y = 0, 1, \dots$ . Sea  $\log \mu = \theta$ ,  $b(\theta) = e^\theta$ ,  $\phi = 1$  y  $c(y, \phi) = -\log y!$  obtenemos (4). Por tanto, se deduce que  $V(\mu) = \mu$ . (8)

### Binomial

Sea  $Y^*$  la proporción de aciertos en  $n$  ensayos independientes, cada uno con probabilidad de ocurrencia  $\mu$ . Suponemos que  $nY^* \sim B(n, \mu)$ . La función de probabilidad función de probabilidad de  $Y^*$  se expresa entonces en la forma

$$\binom{n}{ny^*} \mu^{ny^*} (1 - \mu)^{n - ny^*} = \exp\left\{\log \binom{n}{ny^*} + ny^* \log \left(\frac{\mu}{1 - \mu}\right) + n \log(1 - \mu)\right\}$$

donde  $0 < \mu, y^* < 1$ . Obtenemos (4) haciendo  $\phi = n$ ,  $\theta = \log \left\{\frac{\mu}{1 - \mu}\right\}$ ,  $b(\theta) = \log(1 + e^\theta)$  y  $c(y^*, \phi) = \log \left(\frac{\phi}{\phi y^*}\right)$ . La función de varianza viene dada por  $V(\mu) = \mu(1 - \mu)$ . (8)

### Gamma

Sea  $Y$  una variable aleatoria con distribución gamma de media  $\mu$  y coeficiente de variación  $\phi^{-\frac{1}{2}}$ , denotamos  $Y \sim G(\mu, \phi)$ . La función de densidad de  $Y$  viene dada por

$$\frac{1}{\Gamma(\phi)} \left(\frac{y\phi}{\mu}\right)^\phi \exp\left(-\frac{y\phi}{\mu}\right) d(\log y) = \exp\left[\phi \left\{\left(-\frac{y}{\mu}\right) - \log \mu\right\} - \log \Gamma(\phi) + \phi \log(\phi y) - \log y\right]$$

Donde  $y, \phi, \mu > 0$  y  $\Gamma(\phi) = \int_0^\infty t^{\phi-1} e^{-t} dt$  es la función gamma. Luego, haciendo  $\theta = -\frac{1}{\mu}$ ,  $b(\theta) = -\log(-\theta)$ ,  $c(y, \phi) = (\phi - 1) \log y + \phi \log \phi - \log \Gamma(\phi)$  obtenemos (4).

Para  $0 < \phi < 1$  la densidad de rango tiene un polo en el origen y disminuye monótonicamente cuando  $y \rightarrow \infty$ . La exponencial es un caso especial cuando  $\phi = 1$ . Para  $\phi > 1$  la función de densidad asume cero en el origen, tiene un máximo en  $y = \mu - \frac{\mu}{\phi}$  y luego disminuye hasta  $y \rightarrow \infty$ . A  $\chi_k^2$  es otro caso especial cuando  $\phi = \frac{k}{2}$  y  $\mu = k$ . La distribución normal esto se obtiene haciendo  $\phi \rightarrow \infty$ . Es decir, cuando  $\phi$  es grande  $Y \sim N(\mu, \phi^{-1}V(\mu))$ . Tenemos que  $\phi = \frac{E^2(Y)}{\text{Var}(Y)}$  es la inversa del coeficiente de variación de  $Y$  al cuadrado, es decir,  $\phi = \frac{1}{(CV)^2}$  donde  $CV = \frac{\sqrt{\text{Var}(Y)}}{E(Y)}$ . La función de varianza viene dada por  $V(\mu) = \mu^2$ . (8)

### Normal inversa

Sea  $Y$  una variable aleatoria con distribución normal inversa de media  $\mu$  y parámetro de precisión  $\phi$ , denotada por  $Y \sim NI(\mu, \phi)$  y cuya función de densidad viene dada por

$$\frac{\phi^{\frac{1}{2}}}{\sqrt{2\pi y^3}} \exp \left\{ -\frac{\phi(y - \mu)^2}{2\mu^2 y} \right\} = \exp \left[ \phi \left\{ -\frac{y}{2\mu^2} + \frac{1}{\mu} \right\} - \frac{1}{2} \left\{ \log \left( \frac{2\pi y^3}{\phi} \right) + \frac{\phi}{y} \right\} \right]$$

donde  $y > 0, \mu > 0$ . Haciendo  $\theta = -\frac{1}{2\mu^2}$ ,  $b(\theta) = -(-2\theta)^{\frac{1}{2}}$  y  $c(y, \phi) = \frac{1}{2} \log \left\{ \frac{\phi}{2\pi y^3} \right\} - \frac{\phi}{2y}$  obtenemos (4). La normal inversa se acerca a la normal cuando  $\phi \rightarrow \infty$ . Es decir, para grandes  $\phi$  tenemos que  $Y \sim N(\mu, \phi^{-1}V(\mu))$ . A La función de varianza viene dada por  $V(\mu) = \mu^3$ . (8)

Las distribuciones más importantes de la forma (1.4) de las que nos ocuparemos se resumen en la tabla [I](#).

Cuadro 1: Características de algunas distribuciones univariantes comunes en la familia exponencial

	Normal	Poisson	Binomial	Gamma	N. Inversa
Notación	$N(\mu, \sigma^2)$	$P(\mu)$	$B(n, \mu)$	$G(\mu, \phi)$	$NI(\mu, \phi)$
Rango de $y$	$\infty$	$\mathbb{N}$	$< 1$	$> 0$	$> 0$
Parámetro de dispersión: $\phi$	$\sigma^{-2}$	1	$n$	$\frac{1}{(CV)^2}$	$\phi$
Función acumulativa: $b(\theta)$	$\frac{\theta^2}{2}$	$e^\theta$	$\log(1 + e^\theta)$	$-\log(-\theta)$	$-(-2\theta)^{\frac{1}{2}}$
$\mu(\theta) = E(Y; \theta)$	$\theta$	$e^\theta$	$\frac{e^\theta}{1+e^\theta}$	$-\frac{1}{\theta}$	$(-2\theta)^{\frac{1}{2}}$
Enlace canónico: $\theta(\mu)$	identidad	log	logit	reciprocal	$\frac{1}{\mu^2}$
Función de la varianza: $V(\mu)$	1	$\mu$	$\mu(1 - \mu)$	$\mu^2$	$\mu^3$

El parámetro canónico, denotado por  $\theta$ , se define por (4). La relación entre  $\mu$  y  $\theta$  se da en las líneas 5 y 6 de la tabla.

## Funciones de enlace

La función de enlace relaciona el predictor lineal  $\eta$  con el valor esperado  $\mu$  de un dato  $y$ . En los modelos lineales clásicos la media y el predictor lineal son idénticos, y el enlace idéntico es plausible en la medida en que tanto  $\eta$  como  $\mu$  pueden tomar cualquier valor en la recta real. Sin embargo, cuando tratamos con recuentos y la distribución es Poisson, debemos tener  $\mu > 0$ , por lo que el vínculo de identidad es menos atractivo, en parte porque  $\eta$  puede ser negativo mientras que  $\mu$  no debe serlo. Los modelos para recuentos basados en la independencia en datos de clasificación cruzada conducen naturalmente a efectos multiplicativos, y esto se expresa mediante el enlace logarítmico,  $\eta = \log \mu$ , con su inverso  $\mu = e^\eta$ . Ahora los efectos aditivos que contribuyen a  $\eta$  se convierten en efectos multiplicativos que contribuyen a  $\mu$  y  $\mu$  es necesariamente positivo. [\[8\]](#)

Para la distribución binomial tenemos  $0 < \mu < 1$  y un enlace debe satisfacer la condición de que mapea el intervalo  $(0, 1)$  sobre toda la recta real. En lo sucesivo consideraremos tres funciones principales, a saber:

1. logit

$$\eta = \log \left\{ \frac{\mu}{(1 - \mu)} \right\}$$

2. probit

$$\eta = \Phi^{-1}(\mu)$$

donde  $\Phi(\cdot)$  es la función de distribución acumulativa normal;

3. complementario log-log

$$\eta = \log \{-\log(1 - \mu)\}$$

La familia de potencia de los enlaces es importante, al menos para las observaciones con media positiva. Esta familia puede especificarse mediante

$$\eta = \frac{(\mu^\lambda - 1)}{\lambda} \tag{1.11a}$$

con el valor límite

$$\eta = \log \mu; \quad \text{como } \lambda \rightarrow 0 \tag{1.11b}$$

o por

$$\eta = \begin{cases} \mu^\lambda & \lambda \neq 0 \\ \log \mu & \lambda = 0 \end{cases} \tag{1.12}$$

La primera forma tiene la ventaja de una transición suave cuando  $\lambda$  pasa por cero, pero con cualquiera de las dos formas hay que tomar medidas especiales en cualquier cálculo con  $\lambda = 0$ . [\[8\]](#)



## Estadísticas suficientes y enlaces canónicos

Cada una de las distribuciones de la tabla 1 tiene una función de enlace especial para la que existe un estadístico suficiente igual en dimensión a  $\beta$  en el predictor lineal  $\eta = \sum x_j \beta_j$ . Estos enlaces canónicos, como se llamarán, se producen cuando

$$\theta = \eta$$

donde  $\theta$  es el parámetro canónico definido en (1.4) y mostrado en la Tabla 1. Los enlaces canónicos para las distribuciones de dicha tabla son, por tanto, los siguientes:

Cuadro 2: Los enlaces canónicos para las distribuciones en la familia exponencial

Normal	$\eta = \mu,$
Poisson	$\eta = \log \mu,$
Binomial	$\eta = \log \left\{ \frac{\mu}{1-\mu} \right\},$
Gamma	$\eta = \mu^{-1},$
N. Inversa	$\eta = \mu^{-2},$

Para los enlaces canónicos, la estadística suficiente es  $X^T Y$  en notación vectorial, con componentes

$$\sum_i x_{ij} Y_i, \quad j = 1, 2, \dots, p$$

sumando sobre las unidades. Sin embargo, hay que tener en cuenta que, aunque los enlaces canónicos conducen a propiedades estadísticas deseables del modelo, especialmente en muestras pequeñas, en general no hay ninguna razón a priori para que los efectos sistemáticos de un modelo sean aditivos en la escala dada por ese enlace. Es conveniente que lo sean, pero la conveniencia por sí sola no debe sustituir a la calidad del ajuste como criterio de selección del modelo. Sin embargo, descubriremos que los vínculos canónicos son a menudo eminentemente sensatos desde el punto de vista científico.

### Otros enlaces

#### Enlace Probit

Sea  $\mu$  la proporción de aciertos de una distribución binomial. El enlace se define por

$$\Phi^{-1}(\mu) = \eta,$$

donde  $\Phi(\cdot)$  es la función de distribución acumulativa de la normal estándar.

### Enlace complementaria log-log

La distribución de valores extremos (logaritmo de la exponencial) tiene una función de densidad dada por

$$f(y) = \exp \{y - \exp(y)\}$$

donde  $-\infty < y < \infty$ . Entonces, la función de distribución acumulativa viene dada por

$$f(y) = 1 - \exp \{-\exp(y)\}$$

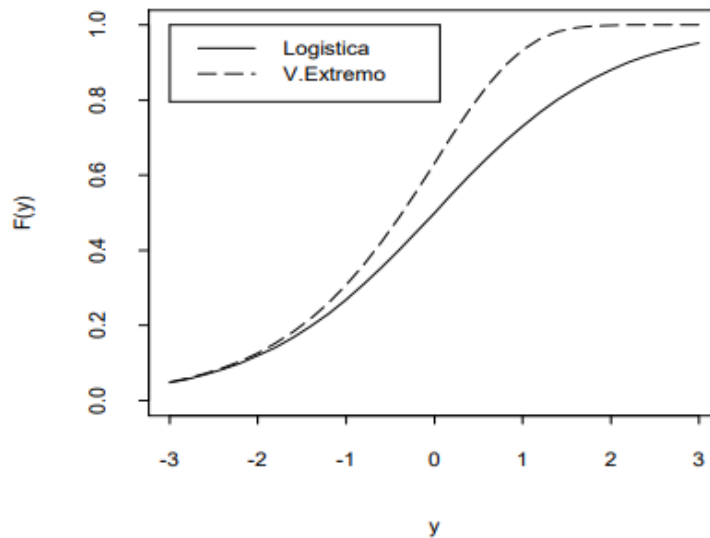


Figura 1: Función de distribución acumulada de las curvas logísticas y de valores extremos.

El modelo binomial con enlace complementario log-log se define como

$$\mu = 1 - \exp \{-\exp(\eta)\}$$

o, de forma equivalente,

$$\log \{-\log(1 - \mu)\} = \eta$$

El enlace logito se define de forma similar. La función de densidad de la distribución logística viene dada por

$$f(y) = \frac{\exp(y)}{\{1 + \exp(y)\}^2}$$

donde  $-\infty < y < \infty$ . De ello se deduce que la función de distribución acumulativa se expresa en la forma

$$F(y) = \frac{e^y}{(1 + e^y)}$$

El modelo logístico binomial se obtiene sustituyendo  $F(y)$  por  $\mu$  e  $y$  por  $\eta$  en la expresión anterior. Como en el caso binomial el parámetro de interés es siempre una probabilidad, es muy razonable que se utilicen funciones de distribuciones acumulativas para generar nuevas conexiones y, en consecuencia, nuevos modelos. En la figura 1.1 mostramos la  $F(y)$  de la distribución logística y de la distribución de valores extremos para valores de  $y$  que varían en el intervalo  $[-3,3]$ . Tenemos que la curva logística es simétrica en torno a  $F(y) = \frac{1}{2}$ , mientras que la curva de valores extremos muestra comportamientos diferentes para  $F(y) \leq \frac{1}{2}$  y  $F(y) > \frac{1}{2}$ .

### Enlace Box-Cox

Una clase importante de enlaces, al menos para las observaciones positivas, es la clase de enlaces Box-Cox definida por

$$\eta = \frac{\mu^\lambda - 1}{\lambda}$$

para  $\lambda \neq 0$  y  $\eta = \log \mu$  para  $\lambda \rightarrow 0$ . La idea ahora es aplicar la transformación Box-Cox a la media de la variable de respuesta

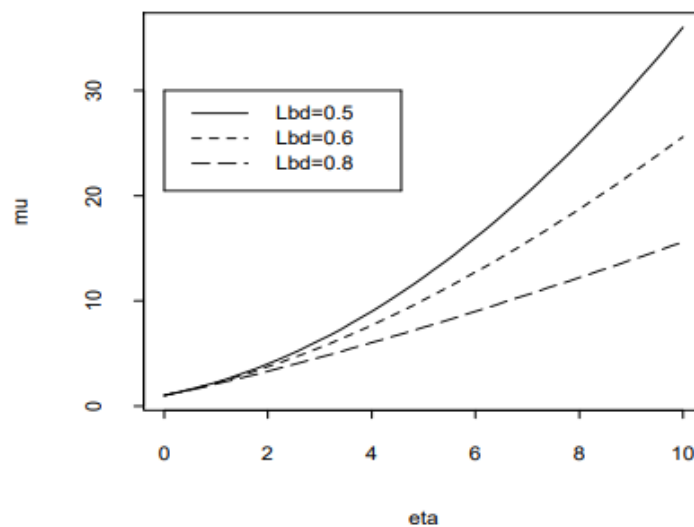


Figura 2: Enlace Box-Cox para algunos valores de  $\lambda$ .

en lugar de transformar la propia variable de respuesta. La figura 1.2 muestra el comportamiento de  $\mu$  para algunos valores de  $\lambda$  y para  $\eta$  que varía en el intervalo  $[0,10]$ .

### Enlace Aranda-Ordaz

Otra transformación importante fue propuesta por Aranda-Ordaz (1981) para datos binarios. La transformación viene dada por

$$\eta = \log \left\{ \frac{(1 - \mu)^{-\alpha} - 1}{\alpha} \right\}$$

donde  $0 < \mu < 1$  y  $\alpha$  es una constante desconocida. Cuando  $\alpha = 1$  tenemos el enlace logit  $\eta = \log \left\{ \frac{\mu}{(1-\mu)} \right\}$ . Cuando  $\alpha \rightarrow 0$  tenemos  $\frac{\{(1-\mu)^{-\alpha}-1\}}{\alpha} \rightarrow \log(1 - \mu)^{-1}$  por lo que  $\eta = \log \{-\log(1 - \mu)\}$  y por tanto obtenemos el enlace

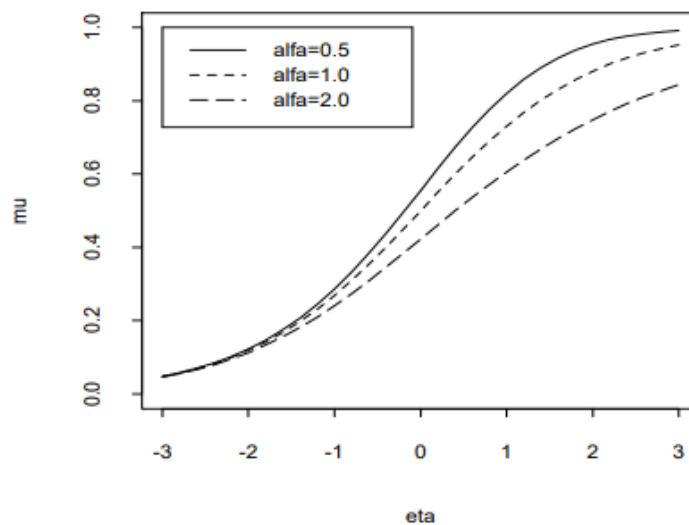


Figura 3: Enlace Aranda-Ordaz para algunos valores de  $\alpha$ .

complemento log-log. En la figura 1.3 vemos el comportamiento de  $\mu$  para algunos valores de  $\alpha$ . En muchas situaciones prácticas el interés puede ser probar si el modelo logístico es apropiado,  $H_0 : \alpha = 1$ , contra la necesidad de una transformación de enlace,  $H_1 : \alpha \neq 1$ .

Los GLM se ajustan en la aplicación R utilizando el comando *glm*. Para ilustrar una aplicación, supongamos que estamos interesados en ajustar un modelo de Poisson con enlace canónico y que la variable de respuesta se denota por *resp* con variables explicativas *cov1* y *cov2*. Podemos enviar los resultados del ajuste a un archivo (objeto en R), por ejemplo con nombre *fit.poisson*, utilizando el comando

```
fit.poisson = glm(resp ~ cov1 + cov2, family = poisson)
```

Con el comando

```
summary(fit.poisson)
```

podemos obtener un resumen de los resultados del ajuste.

## Principios para reducir los datos

El uso de cualquier estadístico  $T(X)$  implica una reducción de los datos muestrales. Sea  $X = (X_1, \dots, X_n)$  una muestra aleatoria simple (un vector aleatorio) y sea  $x = (x_1, \dots, x_n)$ ,  $y = (y_1, \dots, y_n)$  muestras observadas (realizaciones de  $X$ ). Si decidimos usar el estadístico  $T(X)$  en vez de toda la muestra, serán tratadas igual dos muestras observadas cualesquiera  $x$ ,  $y$ , siempre que  $T(x) = T(y)$ . Es decir, al usar el estadístico  $T$ , en lugar de toda la muestra, se pierde información. Se plantea así el problema de buscar estadísticos  $T$  tales que la información que se pierde al usarlos sea irrelevante para los fines que nos hayamos marcado. (5)

Dado el espacio muestral  $\chi$ , la imagen de  $\Omega$  mediante el vector aleatorio  $X$ , reducir los datos en términos de un estadístico  $T$  es equivalente a dar una partición de  $\chi$ . En efecto, sea  $T = \{t : t = T(x), \text{ para algún } x \in \chi\}$ , la imagen de  $\chi$  mediante el estadístico  $T$ . Entonces  $\{A_t = T^{-1}(t) : t \in T\}$  es una partición de  $\chi$  inducida por  $T$ . Al observar  $x$  y limitarnos a registrar el valor  $T(x)$ , podremos saber que hemos observado un elemento de  $A_{T(x)}$ , pero desconoceremos cuál de ellos. (5)

### Principio de suficiencia

Un estadístico  $T$  es suficiente para un parámetro  $\theta$  si captura toda la información que sobre  $\theta$  contiene la muestra. Cualquier información adicional (es decir, aparte del valor del estadístico  $T$  que la muestra pueda aportar, no proporciona información relevante sobre  $\theta$ . Estas consideraciones se concretan en el siguiente principio:

**Definición 2** (Principio de suficiencia). Si  $T$  es un estadístico suficiente para  $\theta$ , cualquier inferencia sobre  $\theta$  ha de depender de la muestra  $X = (X_1, \dots, X_n)$  sólo a través del valor  $T(X)$ . Es decir, si  $x$  e  $y$  son tales que  $T(x) = T(y)$ , entonces la inferencia que se haga sobre  $\theta$  será la misma tanto si se observa  $x$  como si se observa  $y$ . (5)

Formalmente, diremos que un estadístico  $T$  es SUFICIENTE para  $\theta$  si la distribución condicionada de  $X$  dado el valor  $T(X)$ , no depende de  $\theta$ . (5)

Veamos, en el caso discreto, que la información que sobre un parámetro aporta un estadístico suficiente es toda la información que aportaría la muestra completa. En primer lugar, si  $t$  es uno de los posibles valores de  $T(X)$ , es decir, si  $P_\theta(T(X) = t) > 0$ , entonces

$$P_\theta(X = x | T(X) = t) = \begin{cases} P_\theta(X = x | T(X) = T(x)) & \text{si } T(x) = t \\ 0 & \text{si } T(x) \neq t \end{cases}$$

Así que sólo son de interés las probabilidades condicionadas  $P_\theta(X = x|T(X) = T(x))$ . Si  $T$  es suficiente estas probabilidades no depende de  $\theta$ , luego,

$$P_\theta(X = x|T(X) = T(x)) = P(X = x|T(X) = T(x))$$

Para todo  $\theta$ . En este sentido entendemos que  $T$  captura toda la información sobre  $\theta$ . (5)

**Teorema 1.** Si  $f(x|\theta)$  es la verosimilitud de un vector aleatorio  $X$  y  $q(t|\theta)$  es la verosimilitud (función de densidad o de masa) de un estadístico  $T(X)$ , se tiene la siguiente equivalencia.  $T(X)$  es un estadístico suficiente para  $\theta$  si y sólo si para cada  $x$  del espacio muestral  $\chi$  el cociente

$$\frac{f(x|\theta)}{q(T(x)|\theta)}$$

no depende de  $\theta$ .

*Demostración.* Haremos la prueba para el caso discreto. El caso continuo no presenta dificultades adicionales.

Veamos primero que  $\frac{f(x|\theta)}{q(T(x)|\theta)}$  no depende de  $\theta$  entonces  $T$  es suficiente. Como ya se vio anteriormente,

$$P_\theta(X = x|T(X) = t) = \begin{cases} P_\theta(X = x|T(X) = T(x)) & \text{si } T(x) = t \\ 0 & \text{si } T(x) \neq t \end{cases}$$

Luego si  $T(X) \neq t$  la función de masa de probabilidad condicionada vale 0 y, no depende de  $\theta$ . En el otro caso,

$$P_\theta(X = x|T(X) = T(x)) = \frac{P_\theta(X = x, T(X) = T(x))}{P_\theta(T(X) = T(x))} = \frac{P_\theta(X = x)}{P_\theta(T(X) = T(x))} = \frac{f(x|\theta)}{q(T(x)|\theta)}$$

que tampoco depende de  $\theta$ , de donde se sigue que  $T$  es suficiente.

Para probar el recíproco, basta con leer la igualdad anterior en orden inverso,

$$\frac{f(x|\theta)}{q(T(x)|\theta)} = P_\theta(X = x|T(X) = T(x))$$

que no depende de  $\theta$  si  $T$  es suficiente. (5) □

El teorema 1 es útil para comprobar si un estadístico es suficiente, pero no lo es para buscar estadístico suficiente. El siguiente teorema sí podrá usarse para este cometido: permite encontrar estadístico suficiente mediante la inspección de la función de densidad o la función de masa de probabilidad, según el caso. (5)

**Teorema 2** (Teorema de Factorización). Sea  $f(x|\theta)$  la verosimilitud de  $X = (X_1, \dots, X_n)$ . El estadístico  $T(X)$  es suficiente para  $\theta$  si y solo si existen funciones  $g(t|\theta)$  y  $h(x)$  tales que para cualquier  $x \in \chi$  y cualquier  $\theta \in \Theta$  la función  $f(x|\theta)$  puede factorizarse así:

$$f(x|\theta) = g(T(x)|\theta)h(x).$$

*Demostración.* (En el caso discreto) Supongamos primero que  $T$  es suficiente. Tomemos  $g(t|\theta) = P_\theta(T(X) = t)$  y  $h(x) = P(X = x|T(X) = T(x))$ , que no depende de  $\theta$  al ser  $T$  suficiente. Con esta elección,

$$\begin{aligned} f(x|\theta) &= P_\theta(X = x) \\ &= P_\theta(X = x \text{ y } T(X) = T(x)) \\ &= P_\theta(T(X) = T(x))P(X = x|T(X) = T(x)) \\ &= g(T(x)|\theta)h(x). \end{aligned}$$

Veamos ahora el recíproco. Supongamos que  $f(x|\theta) = g(T(x)|\theta)h(x)$  y que  $q(x|\theta)$  es la función de masa de probabilidad de  $T(X)$ . Entonces,

$$\begin{aligned} \frac{f(x|\theta)}{q(T(x)|\theta)} &= \frac{g(T(x)|\theta)h(x)}{q(T(x)|\theta)} \\ &= \frac{g(T(x)|\theta)h(x)}{\sum_{A_{T(x)}} g(T(y)|\theta)h(y)} \\ &= \frac{g(T(x)|\theta)h(x)}{g(T(x)|\theta) \sum_{A_{T(x)}} h(y)} \\ &= \frac{h(x)}{\sum_{A_{T(x)}} h(y)}. \end{aligned}$$

que es independiente de  $\theta$ . Luego  $T$  es suficiente para  $\theta$ . (5) □

Según el teorema 2, para encontrar un estadístico suficiente se factoriza la función de densidad (o la de masa) en dos partes: una,  $h(x)$ , que no contenga al parámetro  $\theta$ , y la otra que dependa de la muestra sólo a través de una función suya  $T(x)$ . El estadístico suficiente será entonces  $T(X)$ . (5)

### Estadísticos suficientes r-dimensional

Hasta ahora hemos visto únicamente estadísticos suficientes con valores reales, es decir, unidimensionales. Sin embargo, en la aplicación del teorema de factorización puede ocurrir que la función  $g(t|\theta)$  dependa de la muestra a través de más de una función suya. En ese caso la colección de esas funciones, digamos  $T(X) = (T_1(X), \dots, T_r(X))$ , es un estadístico suficiente r-dimensional. Las definiciones y

resultados referentes a la suficiencia ya visto para el caso unidimensional también se aplica al caso de r-dimensional. Usualmente, si el parámetro  $\theta$  es de dimensión mayor que 1, entonces también lo será el estadístico suficiente. (5)

### Estadísticos suficientes minimales

La factorización de la función de verosimilitud no es única y como consecuencia de ello, tampoco es único el estadístico suficiente para un parámetro. Ya vimos que cualquier transformación biyectiva de un estadístico suficiente da lugar a otro estadístico suficiente. Pero aún hay muchos estadísticos suficientes. Por ejemplo, la muestra completa  $X$  también es estadístico suficiente para el parámetro:

$$f(x|\theta) = g(x|\theta)h(x),$$

donde  $h(x) = 1$ ,  $T(x) = x$  y  $g(x|\theta) = f(x|\theta)$ .

Un estadístico suficiente  $T(X)$  se llama MINIMAL si para cualquier otro estadístico  $S(X)$  se tiene que  $T(X)$  es función de  $S(X)$ . Es decir, si ocurre que  $S(x) = S(y)$  entonces forzosamente se tiene que  $T(x) = T(y)$ . Otra forma de expresar esto es la siguiente: sea  $\{A_t : t \in T\}$  la partición del espacio muestral inducido por el estadístico  $T$  y sea  $\{B_s : s \in S\}$  la partición inducida por  $S$ ; para cada  $s \in S$  existe un  $t \in T$  tal que  $B_s \subseteq A_t$ . La partición inducida por cualquier estadístico suficiente es mas fina que la inducida por el estadístico suficiente minimal. (5)

El siguiente teorema proporciona un método para encontrar el estadístico minimal.

**Teorema 3.** *Sea  $f(x|\theta)$  la función de verosimilitud conjunta de  $X$  (discreta o continua). Supongamos que existe una función  $T(x)$  tal que para cualquier par de elementos del espacio muestral  $x, y$ , el cociente*

$$\frac{f(x|\theta)}{f(y|\theta)}$$

*es constante como función de  $\theta$ , si y sólo si  $T(x) = T(y)$ . Entonces  $T(X)$  es un estadístico suficiente minimal para  $\theta$ . (5)*

*Demostración.* Para simplificar la demostración, supongamos  $f(x|\theta) > 0$  para todo  $x \in \chi$  y  $\theta$

Primero mostramos que  $T(X)$  es una estadística suficiente. Sea  $T = \{t : t = T(x) \text{ para algunos } x \in \chi\}$  sea la imagen de  $\chi$  bajo  $T(x)$ . Definir los conjuntos de partición inducidos por  $T(x)$  como  $A_t = \{x : T(x) = t\}$ . Para cada  $A_t$ , elegir y fijar un elemento  $x_t \in A_t$ . Para cualquier  $x \in \chi$ ,  $x_{T(x)}$  es el elemento fijo que está en el mismo conjunto,  $A_t$ , como  $x$ . Ya que  $x$  y  $x_{T(x)}$  están en el mismo



conjunto  $A_t, T(x) = T(x_{T(x)})$  y, por lo tanto,  $\frac{f(x|\theta)}{f(x_{T(x)}|\theta)}$  es constante en función de  $\theta$ . Por lo tanto, podemos definir una función en  $\chi$  por  $h(x) = \frac{f(x|\theta)}{f(x_{T(x)}|\theta)}$  y  $h$  no depende de  $\theta$ . Definir una función en  $T$  por  $g(t|\theta) = f(x_t|\theta)$ . Entonces se puede ver que

$$f(x|\theta) = \frac{f(x_{T(x)}|\theta)f(x|\theta)}{f(x_{T(x)}|\theta)} = g(T(x)|\theta)h(x)$$

y, por el teorema de factorización,  $T(X)$  es una estadística suficiente para  $\theta$ . Ahora para demostrar que  $T(X)$  es minimal, dejemos que  $T'(X)$  ser cualquier otra estadística suficiente. Por el Teorema de la factorización, existen funciones  $g'$  y  $h'$  tales que  $f(x|\theta) = g'(T'(x)|\theta)h'(x)$  Sean  $x$  e  $y$  dos puntos cualesquiera de la muestra con  $T'(x) = T'(y)$ . Entonces

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{g'(T'(x)|\theta)h'(x)}{g'(T'(y)|\theta)h'(y)} = \frac{h'(x)}{h'(y)}$$

Como esta relación no depende de  $\theta$ , las hipótesis del teorema implican que  $T(x) = T(y)$ . Por tanto,  $T(x)$  es una función de  $T'(x)$  y  $T(x)$  es minimal. (5)  $\square$

### Estadísticos ancilares

Se define un ESTADÍSTICO ANCILAR como aquel cuya distribución no depende del parámetro  $\theta$ . Los estadísticos ancilares son los que resumen la parte de información de la muestra que no recogen los estadísticos minimales suficientes. A primera vista, parece que no deben aportar información relevante en la estimación del parámetro de interés  $\theta$ . Sin embargo, hay ocasiones en que en combinación con otros estadísticos si son útiles en la estimación de  $\theta$ . (5)

### Estadísticos completos

Sea  $f_T(t|\theta)$  la función de densidad (o de masa de probabilidad) de un estadístico  $T$ . Diremos que la familia de distribución  $\{f_T(t|\theta) : \theta \in \Theta\}$  es COMPLETA si se da la implicación siguiente:

$$E_\theta(g(T)) = 0 \quad \text{para todo } \theta \rightarrow P_\theta(g(T) = 0) = 1 \quad \text{para todo } \theta$$

En ese caso diremos que  $T$  es un ESTADÍSTICO COMPLETO. (5)

La definición de completitud refuerza la de suficiencia en el sentido de que, si un estadístico es suficiente y completo entonces, por una parte, es suficiente minimal (el recíproco no es cierto) y, por otra, ese estadístico es independiente de cualquier estadístico ancilar. Es decir, en el caso de que exista un estadístico suficiente y completo si es cierta la idea intuitiva de que los estadísticos ancilares no puede

aportar nada relevante a la estimación del parámetro. Esto no ocurre si y sólo si se tiene estadísticos suficientes y minimales. (5)

Los siguientes resultados reflejan formalmente esa idea. (5)

**Teorema 4.** *Si  $T$  es un estadístico suficiente y completo para  $\theta$ , entonces  $T$  es suficiente minimal*

*Demostración.* (Extraída de Cristóbal 1992, p. 200.) Sea  $S$  un estadístico suficiente para  $\theta$ . Entonces, por la ley de la esperanza iterada,

$$E_{\theta} \left( T - E \left( \frac{T}{S} \right) \right) = E_{\theta}(T) - E_{\theta}(T) = 0, \quad \text{para todo } \theta \in \Theta$$

Por ser  $T$  completo, se tiene que

$$P_{\theta} \left( T - E \left( \frac{T}{S} \right) = 0 \right) = 1 \rightarrow T = E \left( \frac{T}{S} \right)$$

casi seguro, luego  $T$  es función de cualquier estadístico suficiente  $S$  y, por lo tanto, es minimal. (5) □

## Probabilidad

El objetivo a continuación es asignar a cada evento  $E$  un número real no negativo que indique el chance que tiene  $E$  de ocurrir. Supóngase que se realice un experimento  $n$  veces y que las condiciones en que éste se ejecuta se mantienen más o menos constantes. (2)

## Axiomas de la Probabilidad

G. Bohlmann él define la probabilidad de un evento  $E$  como un número no negativo  $p(E)$  para el cual se satisface:

- Si  $E$  es un evento seguro entonces  $p(E) = 1$ .
- Si  $E_1$  y  $E_2$  son dos eventos, tales que ellos ocurren simultáneamente con probabilidad cero, entonces la probabilidad de que  $E_1$  o  $E_2$  ocurran es igual a  $p(E_1) + p(E_2)$ . (2)

# Estadística Bayesiana

Una forma de recoger la información previa sobre el parámetro de interés es definir una distribución de probabilidad sobre el parámetro, que se llama distribución a priori, de forma que las regiones más probables a priori sean aquellas que contienen los valores más plausibles según la información previa existente, antes de observar ningún valor de datos. (Bernardo, 1993)

## Teorema de Bayes

Si  $A$  y  $E$  son eventos tales que  $P(E) \neq 0$ ,  $P(A|E)$  y  $P(E|A)$  están relacionados por

$$\begin{aligned} P(A|E) &= \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|A^c)P(A^c)} \\ &= \frac{P(E|A)P(A)}{P(E)} \end{aligned}$$

En particular,

$$\frac{P(A|E)}{P(B|E)} = \frac{P(E|A)}{P(E|B)}$$

cuando  $P(B) = P(A)$ . Derivar este resultado a través de la maquinaria de la moderna teoría de la probabilidad axiomatizada es trivial. Sin embargo, aparece como un importante paso conceptual en la historia de la Estadística, siendo la primera inversión de probabilidades. La ecuación  $\frac{P(A|E)}{P(B|E)} = \frac{P(E|A)}{P(E|B)}$  expresa el hecho fundamental de que, para dos causas equiprobables, el cociente de sus probabilidades dado un efecto particular es el mismo que el cociente de las probabilidades de este efecto dado las dos causas. Este teorema también es un principio de actualización, ya que describe la actualización de la probabilidad de  $A$  de  $P(A)$  a  $P(A|E)$  una vez que se ha observado  $E$ . Thomas Bayes (1764) demostró en realidad una versión continua de este resultado, a saber, que dadas dos variables aleatorias  $x$  e  $y$ , con distribución condicional  $f(x|y)$  y distribución marginal  $g(y)$ , la distribución condicional de  $y$  dada  $x$  es

$$g(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}$$

Aunque este teorema de inversión es bastante natural desde un punto de vista probabilístico Bayes y Laplace fueron más allá y consideraron que la incertidumbre sobre los parámetros  $\theta$  de un modelo podría modelarse a través de una distribución de probabilidad de probabilidad  $\pi$  sobre  $\Theta$ , llamada distribución a priori. La inferencia se basa entonces en la distribución de  $\theta$  condicionada a  $x$ ,  $\pi(\theta|x)$ , llamada

distribución posterior y definida por

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$$

Obsérvese que  $\pi(\theta|x)$  es en realidad proporcional a la distribución de  $x$  condicionada a  $\theta$ , es decir, la probabilidad, multiplicada por la distribución a priori de  $\theta$ . (Parece que la generalidad completa de  $\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$  no fue percibida por Bayes sino por Laplace, que la desarrolló en mayor medida). La principal adición aportado por un modelo estadístico bayesiano es, pues, considerar una distribución de probabilidad sobre los parámetros. (10)

**Definición 3.** Un modelo estadístico bayesiano está formado por un modelo estadístico paramétrico,  $f(x|\theta)$ , y una distribución a priori sobre los parámetros,  $\pi(\theta)$ . (10)

## Distribución a priori y posteriori

Supongamos en este punto que, además de la distribución muestral,  $f(x|\theta)$ , se dispone de una distribución a priori sobre  $\theta$ ,  $\pi(\theta)$ , es decir, que tratamos con un modelo bayesiano completo. Dadas estas dos distribuciones, podemos construir varias distribuciones, a saber (10)

1. la distribución conjunta de  $(\theta, x)$ ,

$$\phi(\theta, x) = f(x|\theta)\pi(\theta)$$

2. la distribución marginal de  $x$ ,

$$\begin{aligned} m(x) &= \int \phi(\theta, x)d\theta \\ &= \int f(x|\theta)\pi(\theta)d\theta \end{aligned}$$

3. la distribución posterior de  $\theta$ , obtenida por la fórmula de Bayes,

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{f(x|\theta)\pi(\theta)}{m(x)} \end{aligned}$$

4. la distribución predictiva de  $y$ , cuando  $y \sim g(y|\theta, x)$ , obtenida por

$$g(y|x) = \int g(y|\theta, x)\pi(\theta|x)d\theta$$

## Regiones de confianza

Además de proporcionar a un decisor aproximaciones del valor verdadero de un parámetro  $\theta$ , es decir, estimadores puntuales, y respuestas a preguntas sobre la inclusión de  $\theta$  en un dominio específico, es decir, procedimientos de prueba, a veces es necesario dar además regiones de confianza sobre  $\theta$ , es decir, subconjuntos  $C_x$  del espacio de parámetros  $\Theta$  en los que  $\theta$  debería estar con alta probabilidad (en el sentido frecuentista o en el bayesiano). Esta noción también se extiende a las transformaciones no biyectivas de  $\theta$ . También es de considerable interés en entornos de previsión y predicción (10)

### 1. Intervalos creíbles

Al igual que en el entorno de las pruebas, el paradigma bayesiano propone una noción de regiones de confianza que es más natural que su contraparte frecuentista, ya que, de nuevo, la notación  $P(\theta \in C_x)$  tiene sentido incluso condicionado a  $x$ . (10)

**Definición 4.** Para una distribución a priori  $\pi$ , se dice que un conjunto  $C_x$  es un conjunto  $\alpha$ -creíble si

$$P^\pi(\theta \in C_x | x) \geq 1 - \alpha$$

Esta región se denomina región  $\alpha$ -creíble HPD (para la densidad posterior más alta) si se puede escribir bajo la forma

$$\{\theta; \pi(\theta|x) > k_\alpha\} \subset C_x^\pi \subset \{\theta; \pi(\theta|x) \geq k_\alpha\},$$

donde  $k_\alpha$  es el mayor límite tal que

$$P^\pi(\theta \in C_x^\alpha | x) \geq 1 - \alpha$$

Considerar sólo las regiones HPD está motivado por el hecho de que minimizan el volumen entre las regiones  $\alpha$ -creíbles y, por lo tanto, pueden ser consideradas como soluciones óptimas en un entorno de decisión. (10)

### 2. Intervalos de confianza clásicos

En la teoría de Neyman–Pearson, las regiones de confianza pueden deducirse de pruebas UMPU mediante un argumento de dualidad: Si

$$C_\theta = \{x; \varphi_\theta(x) = 1\}$$

es la región de aceptación de la hipótesis nula  $H_0 : \theta = \theta_0$ , siendo  $\varphi_{\theta_0}$  una prueba UMPU al nivel  $\alpha$ , la región de confianza correspondiente es

$$\begin{aligned} C_x &= \{\theta; x \in C_\theta\} \\ &= \{\theta; \varphi_\theta(x) = 1\} \end{aligned}$$

y  $P(\theta \in C_x) = 1 - \alpha$ . De forma más general, se dice que una región  $C_x$  es una región de confianza en el nivel  $\alpha$  (en el sentido frecuentista) si, para cada  $\theta \in \Theta$ ,  $P(\theta \in C_x) \geq 1 - \alpha$ . (10)

### 3. Evaluación teórica de conjuntos de confianza

La elección de las regiones HPD suele estar relacionada con un requisito de minimización de volumen, bajo la restricción de cobertura

$$P(\theta \in C_\alpha | x) \geq 1 - \alpha$$

Varios autores han propuesto distintas derivaciones de las regiones de confianza según un criterio puramente teórico de la decisión. Consideran funciones de pérdida que incorporan a la vez requisitos de volumen y de cobertura. (En cierto modo manera, el enfoque anterior corresponde a una pérdida bidimensional con componentes  $\text{vol}(C)$  y  $1 - I_C(\theta)$ ). Por ejemplo, una versión sencilla de esta perspectiva teórica de la decisión consiste en considerar la combinación lineal

$$L(C, \theta) = \text{vol}(C) + cI_{(\theta \notin C)}$$

lo que conlleva el riesgo

$$R(C, \theta) = E[\text{vol}(C_x)] + cP(\theta \notin C_x)$$

(La constante  $c$  puede estar relacionada con un nivel de confianza concreto) (10)

# Metodología

## Naturaleza de la Investigación

La naturaleza de la investigación es de carácter cuantitativo. la metodología se enmarca en un estudio con enfoque bayesiano y clásico. Para determinar cuál de las dos genera mayores resultados. Estudio, no experimental y corte transversal. Se determina de tipo no experimental, debido a que se estudió la situación sin intervenir como investigadores y corte transversal porque vamos a comparar ingresos de dos años.

## Fuente de Información

La fuente de información principal es la página web datos abierto Colombia, de ellos se obtuvo los ingresos de las empresas en el año 2019 y 2020. datos abiertos Colombia, es una página web de datos abiertos de información pública de naturaleza mixta, dispuestas en formatos que permiten su uso y sin restricciones legales, que se rige ante la ley 1712 del 2014 sobre transparencia y acceso a la información pública nacional. Los datos son compilados por cualquier persona sin autorización para que terceros puedan reutilizarlos de forma libre y sin restricciones.

## Población

La base de datos utilizada tiene un registro anual desde el año 2019 hasta el año 2020, en total son 1664 datos, los cuales están divididos por microempresa, pequeña empresa, mediana empresa y macroempresa.

## Muestra

Para este trabajo, Se tiene 365 datos de la tasa representativa del mercado en el año 2019 y 366 datos respectivamente para el año 2020. Luego se genera un muestreo no

probabilístico de los datos de los ingresos de las empresas colombianas. Dividida en grupos como: microempresa, pequeña empresa, mediana empresa, macroempresa, que contenga la misma cantidad de datos que la TRM del año 2019 y 2020.

## Instrumentos

Para llevar a cabo este trabajo de investigación, se organizaron en la Hoja de Cálculo Excel los 1664 datos seleccionados del año 2019 y los 1664 datos seleccionados del año 2020, correspondientes a los ingresos generados por las empresas colombianas. Luego con ayuda del programa RStudio se obtuvo el análisis descriptivo, algunos gráficos que facilitaron la comprensión del comportamiento de los mismos y para calcular la varianza y la media de cada uno y en promedio de todos los grupos.

## Procesamiento y análisis de la información

Para el procesamiento y análisis de la información se utilizaron los siguientes recursos estadísticos:

### Enfoque clásico

#### Distribución Gamma

Sea  $X_1, \dots, X_n$  es una muestra aleatoria de una distribución  $\text{Ga}(\alpha, \beta)$ , con  $\alpha > 0$  y  $\beta > 0$  desconocidos. La función de verosimilitud de  $\alpha$  y  $\beta$  es

$$L(\alpha, \beta|x) = \Gamma(\alpha)^{-n} \beta^{n\alpha} \left( \prod_{j=1}^n x_j \right)^{\alpha-1} e^{-\beta \sum_{j=1}^n x_j}$$

La log-verosimilitud es

$$l(\alpha, \beta|x) = -n \log \Gamma(\alpha) + n\alpha \log \beta + (\alpha - 1) \sum_{j=1}^n \log x_j - \beta \sum_{j=1}^n x_j$$

Al intentar maximizar analíticamente nos encontramos con el problema de derivar  $\log \Gamma(\alpha)$ . Podemos maximizar numéricamente.



## Enfoque Bayesiano

La distribución gamma de parámetros  $\alpha > 0$  y  $\beta > 0$ ,  $\Gamma(\alpha, \beta)$ , tiene función de densidad

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{(\alpha-1)} e^{-\beta x}, \quad 0 < x < \infty$$

donde

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

es la función gamma.

El parámetro  $\alpha$  se denomina parámetro de forma porque es el que determina el grado de asimetría y apuntamiento de la densidad (por ejemplo, el coeficiente de asimetría de una  $\Gamma(\alpha, \beta)$  es  $\frac{2}{\sqrt{\alpha}}$ ). El parámetro  $\beta$  se denomina parámetro de escala porque donde más influye es en la dispersión de la distribución. Concretamente, si  $X \sim \Gamma(\alpha, \beta)$  se cumple que  $E(X) = \alpha_1 \beta_1$  y  $\text{Var}(X) = \alpha_1 \beta_1^2$ .

Obtenidos al solucionar el sistema de ecuaciones  $2 \times 2$  representados por la forma matemática de la esperanza y la varianza de la distribución Gamma de la siguiente forma:

$$\begin{aligned} E(X) &= \alpha_1 \beta_1 \\ \text{Var}(X) &= \alpha_1 \beta_1^2 \end{aligned}$$

Despejando  $\alpha_1$  de la primera ecuación obtenemos

$$\alpha_1 = \frac{E(X)}{\beta_1}$$

Remplazando en la segunda ecuación por  $\alpha_1$  vamos a obtener

$$\begin{aligned} \text{Var}(X) &= \alpha_1 \beta_1^2 \\ &= \left( \frac{E(X)}{\beta_1} \right) \beta_1^2 \\ &= E(X) \beta_1 \end{aligned}$$

Despejando  $\beta_1$  de la ecuación anterior obtenemos

$$\beta_1 = \frac{\text{Var}(X)}{E(X)}$$

y remplazando en la ecuación de  $\alpha_1$  obtenemos

$$\begin{aligned}
\alpha_1 &= \frac{E(X)}{\beta_1} \\
&= \frac{E(X)}{\frac{\text{Var}(X)}{E(X)}} \\
&= \frac{E(X)^2}{\text{Var}(X)}
\end{aligned}$$

Entonces, la función de verosimilitud es de la forma

$$L(x_i|\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Desde el enfoque bayesiano,  $\alpha$  y  $\beta$  son variables aleatorias, por tanto asignamos una distribución a priori de probabilidad

$$\alpha, \beta \sim \text{Gamma}(\alpha_1, \beta_1) = h(\theta) \quad \text{a priori}$$

luego, la distribución a posteriori es de la forma

$$\begin{aligned}
h(x_1, x_2, \dots, x_n | \alpha_1, \beta_1) &\propto L(x_1|\theta)h(\theta) \\
h(x_i | \alpha_1, \beta_1) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \times \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} x^{\alpha_1-1} e^{-\beta_1 x} \\
&= \frac{(\beta_1 + \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1})^{\alpha_1 + \frac{n}{2}}}{\Gamma(\alpha_1 + \frac{n}{2})} x^{(\alpha_1 + \frac{n}{2})-1} e^{-(\beta_1 + \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1})x}
\end{aligned}$$

Por tanto, la distribución a posteriori es una gamma de parámetros

$$\text{Gamma} \left( \alpha_2 = \alpha_1 + \frac{n}{2}, \beta_2 = \beta_1 + \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \right)$$

# Resultados

Los datos utilizados en el presente análisis corresponden a los ingresos registrados de los grupos de empresas ante la superintendencia financiera de Colombia, los cuales se encuentra disponibles en la página web datos abiertos Colombia (<https://www.datos.gov.co/>). Así mismo, el periodo que se analiza corresponde al año 2019 – 2020, esto con la finalidad de identificar los posible traumatismo o estado de crisis originados por la pandemia del coronavirus (COVID 19).

Para efectos del análisis se realizo la segmentación de la información por tipos de empresas, es decir, de la muestra total de empresas que se recolectaron a través de la pagina web, se realizo la originación de subgrupos por Microempresas, Pequeñas empresas, Medianas empresas y Macroempresas.

Luego de dicha segmentación iniciamos analizando los ingresos asociados a las **MICROEMPRESAS** los cuales se presentan en el Cuadro 3.

Cuadro 3: Descriptiva de los ingresos de las MICROEMPRESAS de Colombia

Variable	Ingresos 2019	Ingresos 2020
Media	18093636	8961102
Mediana	11378825	3847215
Varianza	2.304303e+14	1.196676e+14
Des. Estándar	15179932	10939270
IQR	2.3e+07	10350309
Asimétrica	0.8325142	1.690996
Curtosis	2.428489	5.099902

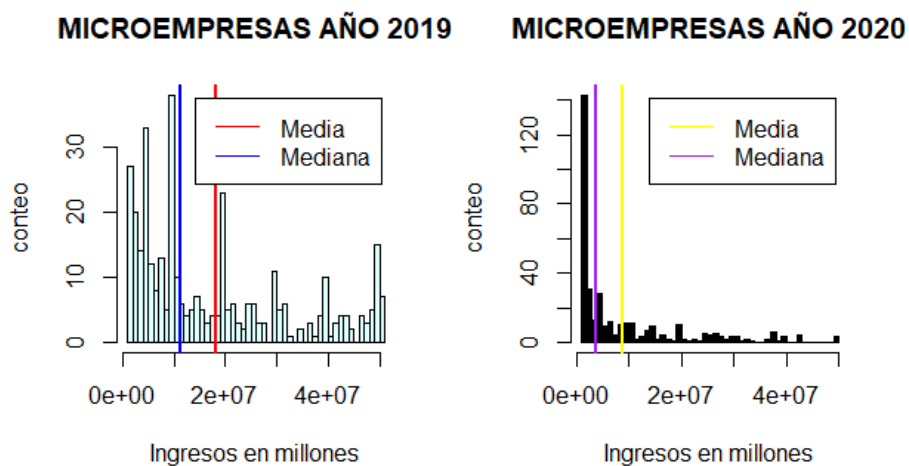
Fuente: Base de datos tomados de la página web datos abiertos Colombia

De acuerdo el Cuadro 3 Para el estudio realizado de las microempresas se observa que los ingresos generados en el periodo 2019 presentaron ganancias en promedio de 18093636 con una dispersión de 15179932 mientras que en el 2020 se generó

en promedio 8961102 y con una dispersión de 10939270. es decir que las microempresas en el año 2020, presentaron una pérdida de 50,5 puntos porcentuales con respecto a los ingresos del año 2019. También se observa que la mitad de las microempresas del 2019 ganan más de 11378825 mientras que en el 2020 tan solo la mitad presentaron ganancias mayores de 3847215, esto presenta un balance crítico para el año 2020 (durante el covid-19). También se observa que ambas presentan una marcada asimetría positiva y leptocúrtica, es decir se logra evidenciar que para ambos periodos presentan datos agrupado que tienden a ingresos bajos. Ahora, en el periodo 2020 presenta menor dispersión de ingresos que el 2019 esto solo refleja que el 50 % de los ingresos están en 230000000 mientras que en el 2020 es de 10350309. Esto quiere decir que la llegada del covid-19 perjudico significativamente los ingresos de las microempresas colombianas.

A continuación, presentamos las correspondientes distribuciones de las variables de estudio, esto es, ingresos de las microempresas para el año 2019 y 2020

Figura 4: Histograma de los ingresos de las MICROEMPRESAS de Colombia 2019 - 2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

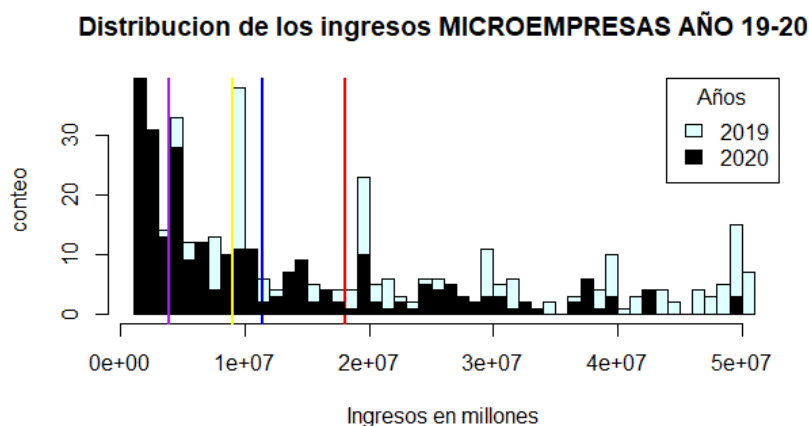
En la gráfica 4, se observa que la gráfica izquierda el comportamiento de los datos están muy relacionados a los presentados en el Cuadro 3, en efecto se evidencia una marcada asimetría positiva en los datos, también podemos observar que la línea roja representa la media para los datos de los ingresos del 2019 que es de 18093636 de lo cual podemos decir que la mayoría de los ingresos de las microempresas estan

por debajo de la media, a diferencia de la línea azul representa la mediana para los datos de los ingresos del 2019 es decir que la mitad de las microempresas generan ingresos menores a 11378825.

De igual manera, en la gráfica derecha se observa el comportamiento de los datos es heterogenio, porque algunas empresas generan ingresos bajos y el resto de las empresas generan ingresos altos, también se observa que la línea amarilla representa la media para los datos de los ingresos del 2020 es decir que las empresas generan en promedio ingresos de 8961102, a diferencia de la línea morada representa la mediana para los datos de los ingresos del 2020 es decir que la mitad de las microempresas generan ingresos menores a 3847215.

A continuación, vamos a mirar la gráfica de los ingresos de las MICROEMPRESAS 2019-2020. En conclusión, al realizar el análisis de los indicadores de tendencia central es claro la robustez de la mediana frente a presencia de datos extremos o asimétricos

Figura 5: Comparacion de los histogramas de los ingresos de las MICROEMPRESAS colombiana 2019–2020

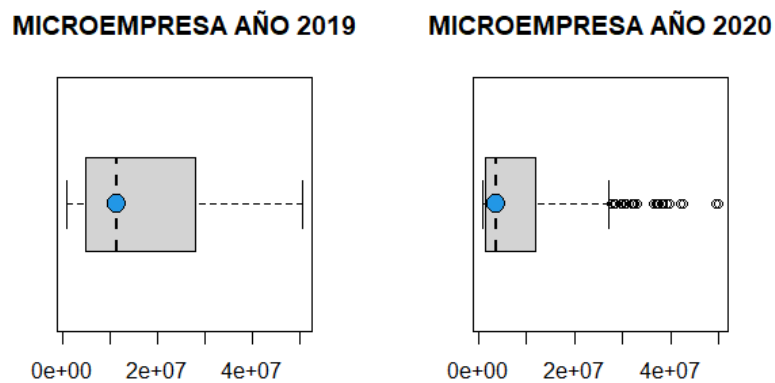


Fuente: Base de datos tomados de la página web datos abiertos Colombia

En la gráfica 5, se observa lo siguiente hay una disminución en los ingresos de las empresas para el año 2020, eso se puede deducir ya que para ese año 2020 muchas empresas bajaron sus ingresos, por las medidas tomadas del gobierno COLOMBIANO. A diferencia del año 2019, teniendo en cuenta que hay una diferencia entre las medidas de tendencia central, la de color rojo y amarillo se hace referencia a la media respectivamente con los años comparados de igual forma el color azul y

morado se hace referencia a la mediana respectivamente con los años comparados. A continuación, vamos a ver la siguiente grafica

Figura 6: Boxplot de los ingresos de las MICROEMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

En la figura 6 se observa una comparación de boxplot de microempresas entre el año 2019 y 2020, lo cual se observa que en el año 2019 los ingresos registrados por las empresas son asimétricos positivos sin ningún valor mayor o menor al rango máximo y mínimo. Pero en el 2020 se observa una asimetría positiva muy sesgada a la derecha, es decir, que muchas empresas generaron ingresos muy por debajo al año anterior y esto hizo que las empresas que no reportaron perdidas o que generaran ganancias se presenten como valores atípicos.

De acuerdo con la metodología presentada para estimar los parámetros de la distribución gamma a través de software estadístico Rstudio encontramos que:

Cuadro 4: Estimaciones de los parámetros con respecto el año 2019 enfoque clásico

	alfa	beta
Momentos	1.424624	7.87362e-08
MV nlm	1.380963	7.87362e-08
MV constrOptim	1.290966	7.137779e-08

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Cuadro 5: Estimaciones de los parámetros con respecto el año 2020 enfoque clásico

	alfa	beta
Momentos	0.6728749	7.508841e-08
MV nlm	0.7510452	7.508841e-08
MV constrOptim	0.8568154	9.565275e-08

Fuente: Base de datos tomados de la página web datos abiertos Colombia

En el Cuadro 4 y 5 se observa el comportamiento de los datos con respecto  $\alpha$  y  $\beta$  con relación al enfoque clásico, mirar en el **Anexo 1**.

Cuadro 6: Estimación de los parámetros enfoque bayesiano

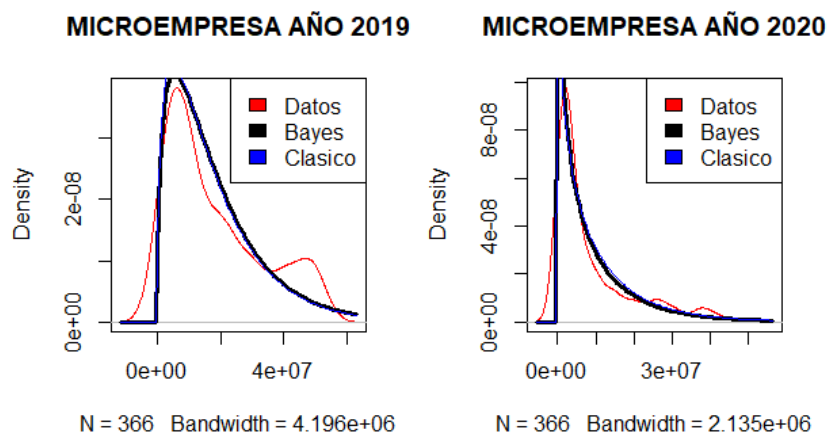
	Ingresos 2019	Ingresos 2020
alfa	1.420732	0.6710365
beta	12735435	13354120

Fuente: Base de datos tomados de la página web datos abiertos Colombia

En el Cuadro 6 se observa el comportamiento de los datos con respecto  $\alpha$  y  $\beta$  con relación al enfoque bayesiano, mirar en el **Anexo 1**.

Teniendo en cuenta los datos arrojados por el software, utilizaremos la función gamma para ver su comportamiento con respecto a la densidad

Figura 7: Gamma (Enfoque Clásico y Enfoque Bayesiano) VS Densidad de los ingresos de las MICROEMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

De la figura 7 se observa dos graficas de las microempresas colombianas del año 2019 y 2020, con enfoques clásico y bayesiano a través de la distribución gamma. Donde se observa que los ingresos en ambos años fueron diferentes. En el año 2019 se presentaron mejores reportes que en el 2020 y además se tiene que el comportamiento de la Gamma es muy similar en ambos enfoques. Por lo tanto, la densidad de la distribución gamma se ajusta a los ingresos de las empresas colombianas tanto en el año 2019 como en el 2020.

Por eso se aplicó un modelo lineal generalizado GLM con respuesta Gamma, a los datos para las MICROEMPRESAS para el año 2019, mirar en el **Anexo 2**. Teniendo en cuenta los datos arrojados por el software, observamos que en la siguiente tabla tenemos:

Cuadro 7: Información del modelo para los ingresos de las MICROEMPRESAS año 2019

Coefficients:				
	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	1.033e-06	6.003e-08	17.21	<2e-16
TRM_2019	-2.910e-10	1.746e-11	-16.67	<2e-16

Fuente: Base de datos tomados de la página web datos abiertos Colombia



Como se observa en el Cuadro 7 en relación de los ingresos 2019 con respecto a la TRM del mismo año, observamos una significancia  $2e - 16$  diferente de 0, por lo cual concluyo que es muy significativo los ingresos del presente año de las MICROEMPRESAS en Colombia con una confiabilidad del 99% y el AIC: 12669.

Del mismo modo se aplicó el modelo, a los datos para las MICROEMPRESAS para el año 2020, mirar en el Anexo 2. Teniendo en cuenta los datos arrojados por el software, observamos que en la siguiente tabla tenemos:

Cuadro 8: Información del modelo para los ingresos de las MICROEMPRESAS año 2020

Coefficients:				
	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	2.341e-08	1.207e-07	0.194	0.846
TRM_2020	2.394e-11	3.280e-11	0.730	0.466

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Como se observa en el Cuadro 8 obtenemos en relación de los ingresos 2020 con respecto a la TRM del mismo año, observamos una significancia 0,846 diferente de 0, por lo cual concluyo que es no es muy significativo los ingresos del presente año de las MICROEMPRESAS en Colombia ya que tenemos una confiabilidad del 15,4% y el AIC: 12456.

Luego de dicha segmentación iniciamos analizando los ingresos asociados a las **PEQUEÑAS EMPRESA** los cuales se presentan en el Cuadro 9.

Cuadro 9: Descriptiva de los ingresos de las PEQUEÑAS EMPRESAS de Colombia

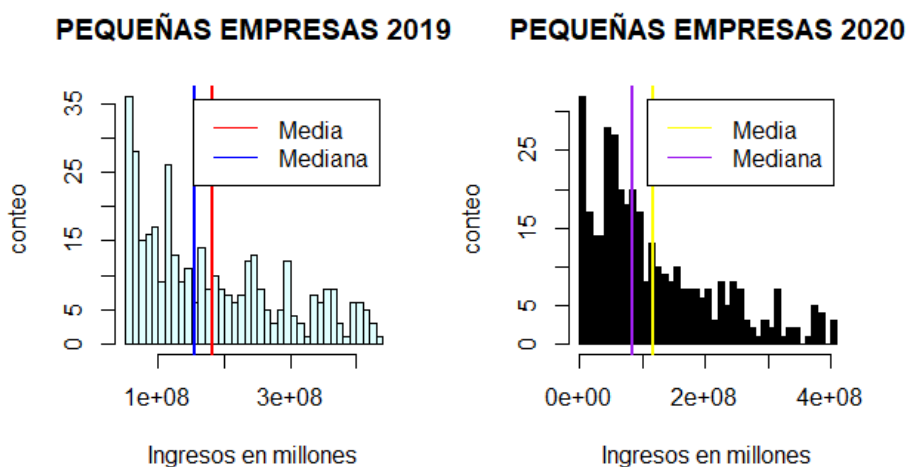
Variable	Ingresos 2019	Ingresos 2020
Media	181838984	117810633
Mediana	153967692	84875000
Varianza	1.169971e+16	9.764435e+15
Des. Estándar	108165178	98815155
IQR	166025142	128336428
Asimétrica	0.6480569	1.03704
Curtosis	2.251094	3.316428

Fuente: Base de datos tomados de la página web datos abiertos Colombia

De acuerdo el Cuadro 9 Para el estudio realizado de las pequeñas empresas se observa que los ingresos generados en el periodo 2019 presentaron ganancias en promedio de 181838984 con una dispersión de 108165178 mientras que en el 2020 se generó en promedio 117810633 y con una dispersión de 98815155, es decir que las pequeñas empresas en el año 2020 presentaron una pérdida de 35 puntos porcentuales con respecto a los ingresos del año 2019. También se observa que la mitad de las pequeñas empresas del 2019 ganan más de 153967692 mientras que en el 2020 tan solo la mitad presentaron ganancias mayores de 98815155, esto presenta un balance crítico para el año 2020 (durante el covid-19). También se observa que ambas presentan una marcada asimetría positiva y leptocúrtica, es decir se logra evidenciar que para ambos periodos presentan datos agrupado que tienden a ingresos bajos. Ahora, en el periodo 2020 presenta menor dispersión de ingresos que el 2019 esto solo refleja que el 50 % de los ingresos están en 166025142 mientras que en el 2020 es de 128336428. Esto quiere decir que la llegada del covid-19 perjudico significativamente los ingresos de las empresas pequeñas en Colombia.

A continuación, presentamos las correspondientes distribuciones de las variables de estudio, esto es, ingresos de las pequeñas empresas para el año 2019 y 2020

Figura 8: Histograma de los ingresos de las PEQUEÑAS EMPRESAS de Colombia 2019 - 2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

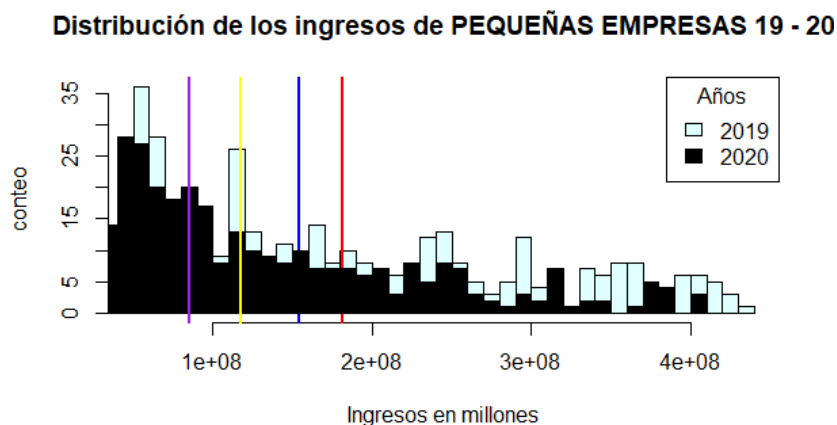
En la gráfica 8, se observa que la gráfica izquierda el comportamiento de los datos están muy relacionados a los presentados en el Cuadro 9, en efecto se evidencia

una marcada asimetría positiva en los datos, también podemos observar que la línea roja representa la media para los datos de los ingresos del 2019 que es de 181838984 de lo cual podemos decir que la mayoría de los ingresos de las pequeñas empresas están por debajo de la media, a diferencia de la línea azul que representa la mediana para los datos de los ingresos del 2019 es decir que la mitad de las pequeñas empresas generan ingresos menores a 153967692.

De igual manera, en la gráfica derecha se observa el comportamiento de los datos es heterogéneo, porque algunas empresas generan ingresos bajos y el resto de las empresas generan ingresos altos, también se observa que la línea amarilla que representa la media para los datos de los ingresos del 2020 es decir que las empresas generan en promedio ingresos de 117810633, a diferencia de la línea morada que representa la mediana para los datos de los ingresos del 2020 es decir que la mitad de las pequeñas empresas generan ingresos menores a 84875000.

A continuación, vamos a mirar la gráfica de los ingresos de las PEQUEÑAS EMPRESAS 2019-2020. En conclusión, al realizar el análisis de los indicadores de tendencia central es claro la robustez de la mediana frente a presencia de datos extremos o asimétricos

Figura 9: Comparación de los histogramas de los ingresos de las PEQUEÑAS EMPRESAS colombiana 2019–2020

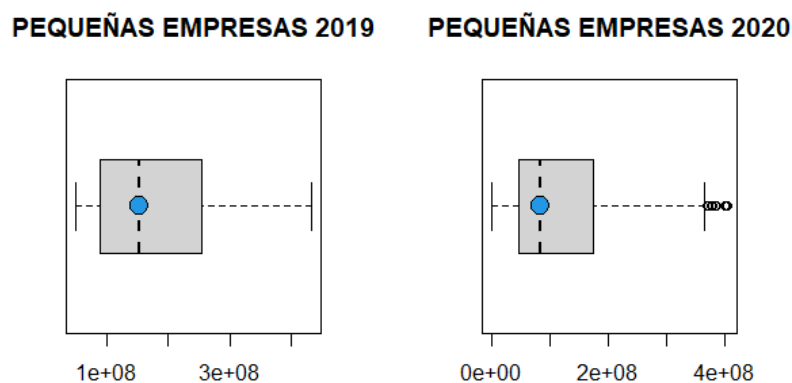


Fuente: Base de datos tomados de la página web datos abiertos Colombia

En la gráfica 9, se observa lo siguiente hay una disminución en los ingresos de las empresas para el año 2020, eso se puede deducir ya que para ese año 2020

muchas empresas bajaron sus ingresos, por las medidas tomadas del gobierno COLOMBIANO. A diferencia del año 2019, teniendo en cuenta que hay una diferencia entre las medidas de tendencia central, la de color rojo y amarillo se hace referencia a la media respectivamente con los años comparados de igual forma el color azul y morado se hace referencia a la mediana respectivamente con los años comparados. A continuación, vamos a ver la siguiente grafica

Figura 10: Boxplot de los ingresos de las PEQUEÑAS EMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

En la figura [10](#) se observa una comparación de boxplot de PEQUEÑAS EMPRESAS entre el año 2019 y 2020, lo cual se observa que en el año 2019 los ingresos registrados por las empresas son asimétricos positivos sin ningún valor mayor o menor al rango máximo y mínimo. Pero en el 2020 se observa una asimetría positiva con algunos ingresos mayores al rango máximo. es decir, que muchas empresas generaron ingresos muy por debajo al año anterior y esto hizo que las empresas que no reportaron perdidas o que generaran ganancias se presenten por fuera del rango máximo.

De acuerdo con la metodología presentada para estimar los parámetros de la distribución gamma a través de software estadístico Rstudio encontramos que:

Cuadro 10: Estimaciones de los parámetros con respecto el año 2019 enfoque clásico

	alfa	beta
Momentos	2.833918	1.558477e-08
MV nlm	2.830429	1.558477e-08
MV constrOptim	2.833918	1.558477e-08

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Cuadro 11: Estimaciones de los parámetros con respecto el año 2020 enfoque clásico

	alfa	beta
Momentos	1.425313	1.209833e-08
MV nlm	1.31273	1.209833e-08
MV constrOptim	1.425313	1.209833e-08

Fuente: Base de datos tomados de la página web datos abiertos Colombia

En el Cuadro [10](#) y [11](#) se observa el comportamiento de los datos con respecto  $\alpha$  y  $\beta$  con relación al enfoque clásico, mirar en el **Anexo 1**.

Cuadro 12: Estimación de los parámetros enfoque bayesiano

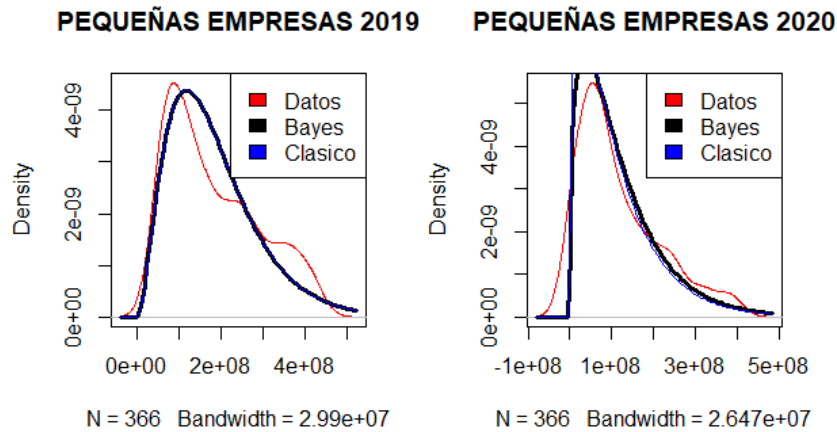
	Ingresos 2019	Ingresos 2020
alfa	2.826175	1.421418
beta	64341020	82882457

Fuente: Base de datos tomados de la página web datos abiertos Colombia

En el Cuadro [12](#) se observa el comportamiento de los datos con respecto  $\alpha$  y  $\beta$  con relación al enfoque bayesiano, mirar en el **Anexo 1**.

Teniendo en cuenta los datos arrojados por el software, utilizaremos la función gamma para ver su comportamiento con respecto a la densidad

Figura 11: Gamma (Enfoque Clásico y Enfoque Bayesiano) VS Densidad de los ingresos de las PEQUEÑAS EMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

De la figura [11](#) se observa dos graficas de las pequeñas empresas colombianas del año 2019 y 2020, con enfoques clásico y bayesiano a través de la distribución gamma. Donde se observa que los ingresos en ambos años fueron diferentes. En el año 2019 se presenta una curva de ingresos bajos y crece hacia ingresos muy altos donde trata de mantenerse bajando muy lento a diferencia del año del 2020 que crece similar al año anterior, pero decae más rápido a ingresos altos. además, se tiene que el comportamiento de la Gamma es muy similar en ambos enfoques. Por lo tanto, la densidad de la distribución gamma se ajusta a los ingresos de las empresas colombianas tanto en el año 2019 como en el 2020.

Por eso se aplicó un modelo lineal generalizado GLM con respuesta Gamma, a los datos para las PEQUEÑAS EMPRESAS para el año 2019, mirar en el **Anexo 2**. Teniendo en cuenta los datos arrojados por el software, observamos que en la siguiente tabla tenemos:

Cuadro 13: Información del modelo para los ingresos de las PEQUEÑAS EMPRESAS año 2019

Coefficients:				
	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	7.480e-08	3.697e-09	20.23	<2e-16
TRM_2019	-2.076e-11	1.090e-12	-19.05	<2e-16

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Como se observa en el Cuadro [13](#) obtenemos en relación de los ingresos 2019 con respecto a la TRM del mismo año, observamos una significancia  $2e - 16$  diferente de 0, por lo cual concluyo que es muy significativo los ingresos del presente año de las PEQUEÑAS EMPRESAS en Colombia con una confiabilidad del 99 % y el AIC: 14138.

Del mismo modo se aplicó el modelo, a los datos para las PEQUEÑAS EMPRESAS para el año 2020, mirar en el **Anexo 2**. Teniendo en cuenta los datos arrojados por el software, observamos que en la siguiente tabla tenemos:

Cuadro 14: Información del modelo para los ingresos de las PEQUEÑAS EMPRESAS año 2020

Coefficients:				
	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	5.029e-09	6.363e-09	0.790	0.430
TRM_2020	9.380e-13	1.725e-12	0.544	0.587

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Como se observa en el Cuadro [14](#) obtenemos en relación de los ingresos 2020 con respecto a la TRM del mismo año, observamos una significancia 0,430 diferente de 0, por lo cual concluyo que no es muy significativo los ingresos del presente año de las PEQUEÑAS EMPRESAS en Colombia ya que tenemos una confiabilidad del 57 % y el AIC: 14343.

Luego de dicha segmentación iniciamos analizando los ingresos asociados a las **MEDIANAS EMPRESAS** los cuales se presentan en el Cuadro [15](#).

Cuadro 15: Descriptiva de los ingresos de las MEDIANAS EMPRESAS de Colombia

Variable	Ingresos 2019	Ingresos 2020
Media	1371435041	1041989034
Mediana	1139982538	801093400
Varianza	6.877244e+17	6.106697e+17
Des. Estándar	829291493	781453591
IQR	1263323676	1043691031
Asimétrica	0.8463248	0.8981396
Curtosis	2.62253	3.085434

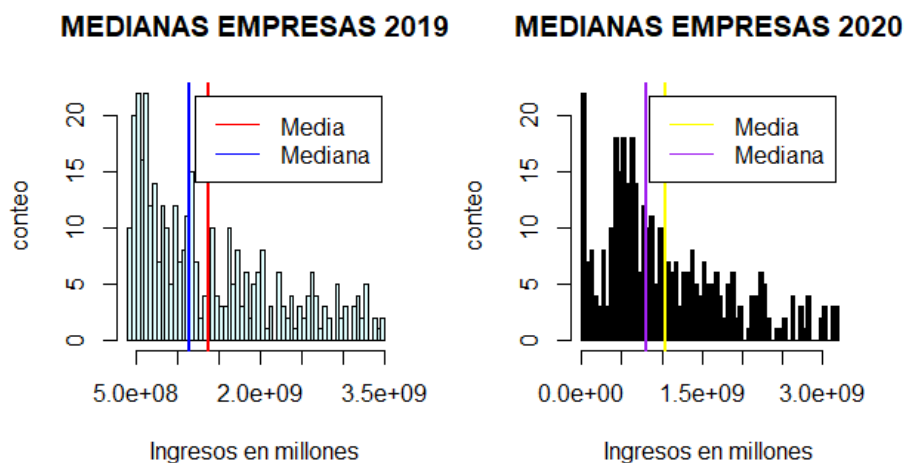
Fuente: Base de datos tomados de la página web datos abiertos Colombia

De acuerdo el Cuadro [15](#) Para el estudio realizado de las medianas empresas se observa que los ingresos generados en el periodo 2019 presentaron ganancias en promedio de 1371435,041 con una dispersión de 829291493 mientras que en el 2020 se generó en promedio 1041989034 y con una dispersión de 781453591, es decir que las medianas empresas en el año 2020 presentaron una pérdida de 24 puntos porcentuales con respecto a los ingresos del año 2019. También se observa que la mitad de las medianas empresas del 2019 ganan más de 1139982538 mientras que en el 2020 tan solo la mitad presentaron ganancias mayores de 801093400, esto presenta un balance crítico para el año 2020 (durante el covid-19). También se observa que ambas presentan una marcada asimetría positiva y leptocúrtica, es decir se logra evidenciar que para ambos periodos presentan datos agrupado que tienden a ingresos bajos. Ahora, en el periodo 2020 presenta menor dispersión de ingresos que el 2019 esto solo refleja que el 50 % de los ingresos están en 1263323676 mientras que en el 2020 es de 1043691031 Esto quiere decir que la llegada del covid-19 perjudico significativamente los ingresos de las empresas medianas en Colombia.

A continuación, presentamos las correspondientes distribuciones de las variables de estudio, esto es, ingresos de las empresas para el año 2019 y 2020



Figura 12: Histograma de los ingresos de las MEDIANAS EMPRESAS de Colombia 2019 - 2020



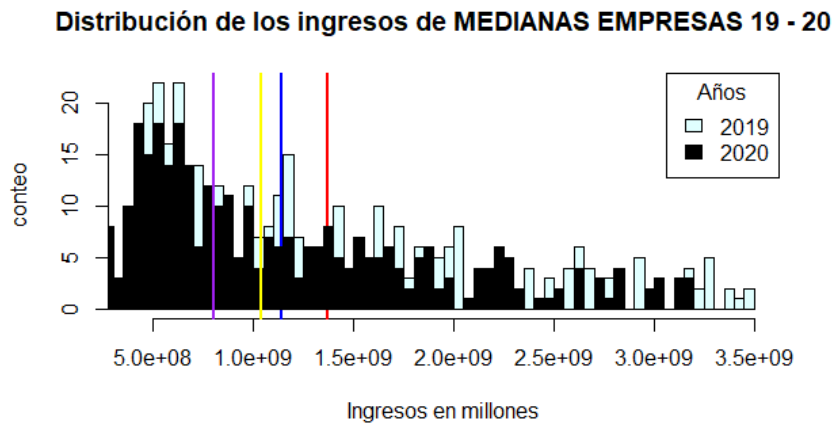
Fuente: Base de datos tomados de la página web datos abiertos Colombia

En la gráfica [12](#), se observa que la gráfica izquierda el comportamiento de los datos están muy relacionados a los presentados en el Cuadro [15](#), en efecto se evidencia una marcada asimetría positiva en los datos, también podemos observar que la línea roja representa la media para los datos de los ingresos del 2019 que es de 1371435041 de lo cual podemos decir que la mayoría de los ingresos de las medianas empresas están por debajo de la media, a diferencia de la línea azul representa la mediana para los datos de los ingresos del 2019 es decir que la mitad de las medianas generan ingresos menores a 1139982538.

De igual manera, en la gráfica derecha se observa el comportamiento de los datos es heterogéneo, porque algunas empresas generan ingresos bajos y el resto de las empresas generan ingresos altos, también se observa que la línea amarilla representa la media para los datos de los ingresos del 2020 es decir que las empresas generan en promedio ingresos de 1041989034, a diferencia de la línea morada representa la mediana para los datos de los ingresos del 2020 es decir que la mitad de las medianas empresas generan ingresos menores a 801093400.

A continuación, vamos a mirar la gráfica de los ingresos de las MEDIANAS EMPRESAS 2019 – 2020. En conclusión, al realizar el análisis de los indicadores de tendencia central es claro la robustez de la mediana frente a presencia de datos extremos o asimétricos

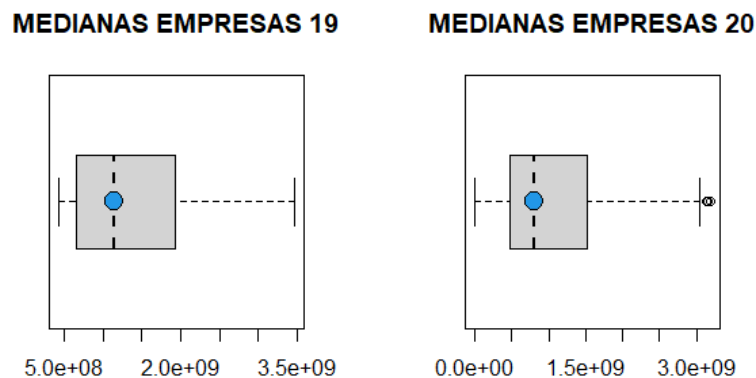
Figura 13: Comparación de los histogramas de los ingresos de las MEDIANAS EMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

En la gráfica [13](#), se observa lo siguiente hay una disminución en los ingresos de las empresas para el año 2020, eso se puede deducir ya que para ese año 2020 muchas empresas bajaron sus ingresos, por las medidas tomadas del gobierno COLOMBIANO. A diferencia del año 2019, teniendo en cuenta que hay una diferencia entre las medidas de tendencia central, la de color rojo y amarillo se hace referencia a la media respectivamente con los años comparados de igual forma el color azul y morado se hace referencia a la mediana respectivamente con los años comparados. A continuación, vamos a ver la siguiente grafica

Figura 14: Boxplot de los ingresos de las MEDIANAS EMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

En la figura 14 se observa una comparación de boxplot de MEDIANAS EMPRESAS entre el año 2019 y 2020, lo cual se observa que en el año 2019 los ingresos registrados por las empresas son asimétricos positivos agrupados a la izquierda sin ningún valor mayor o menor al rango máximo y mínimo. Pero en el 2020 se observa una asimetría positiva muy agrupados a la izquierda con algunos ingresos mayores al rango máximo. es decir, que muchas empresas generaron ingresos muy por debajo al año anterior y esto hizo que las empresas que no reportaron pérdidas o que generaran ganancias se presenten por fuera del rango máximo.

De acuerdo con la metodología presentada para estimar los parámetros de la distribución gamma a través de software estadístico Rstudio encontramos que:

Cuadro 16: Estimaciones de los parámetros con respecto el año 2019 enfoque clásico

	alfa	beta
Momentos	2.742359	1.999627e-09
MV constrOptim	2.818607	2.054883e-09

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Cuadro 17: stimaciones de los parámetros con respecto el año 2020 enfoque clásico

	alfa	beta
Momentos	1.782823	1.71098e-09
MV constrOptim	2.808335	2.047441e-09

Fuente: Base de datos tomados de la página web datos abiertos Colombia

En el Cuadro [16](#) y [17](#) se observa el comportamiento de los datos con respecto  $\alpha$  y  $\beta$  con relación al enfoque clasico, mirar en el **Anexo 1**.

Cuadro 18: Estimación de los parámetros enfoque bayesiano

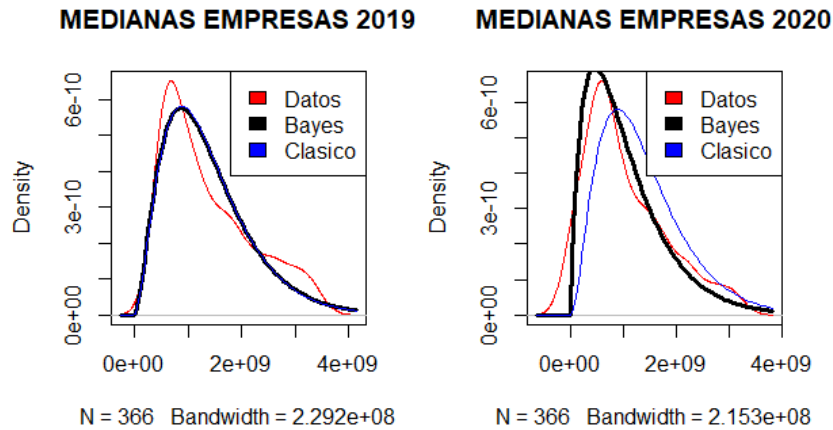
	Ingresos 2019	Ingresos 2020
alfa	2.734866	1.777952
beta	501463328	586061556

Fuente: Base de datos tomados de la página web datos abiertos Colombia

En el Cuadro [18](#) se observa el comportamiento de los datos con respecto  $\alpha$  y  $\beta$  con relación al enfoque bayesiano, mirar en el **Anexo 1**.

Teniendo en cuenta los datos arrojados por el software, utilizaremos la función gamma para ver su comportamiento con respecto a la densidad

Figura 15: Gamma (Enfoque Clásico y Enfoque Bayesiano) VS Densidad de los ingresos de las MEDIANAS EMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

De la figura [15](#) se observa dos graficas de las medianas empresas colombianas del año 2019 y 2020, con enfoques clásico y bayesiano a través de la distribución gamma. Donde se observa que los ingresos no presentan mucha diferencia. En el año 2019 y 2020 presenta una curva de ingresos bajos que crece hacia ingresos altos y va decreciendo lentamente hasta llegar a pocas empresas que generan ingresos muy altos. además, se tiene que el comportamiento de la Gamma en los ingresos de las medianas empresas del año 2019 es muy similar en ambos enfoques, pero en el año 2020 el enfoque bayesiano se ajusta mejor. Por lo tanto, la densidad de la distribución gamma con enfoque bayesiano se ajusta mejor a los ingresos de las empresas colombianas en el año 2019 como en el 2020.

Por eso se aplicó un modelo lineal generalizado GLM con respuesta Gamma, a los datos para las MEDIANA EMPRESAS para el año 2019, mirar en el **Anexo 2**. Teniendo en cuenta los datos arrojados por el software, observamos que en la siguiente tabla tenemos:

Cuadro 19: Información del modelo para los ingresos de las MEDIANAS EMPRESAS año 2019

Coefficients:				
	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	9.736e-09	4.969e-10	19.59	<2e-16
TRM_2019	-2.698e-12	1.465e-13	-18.42	<2e-16

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Como se observa en el Cuadro [19](#) en relación de los ingresos 2019 con respecto a la TRM del mismo año, observamos una significancia  $2e - 16$  diferente de 0, por lo cual concluyo que es muy significativo los ingresos del presente año de las MEDIANAS EMPRESAS en Colombia con una confiabilidad del 99 % y el AIC: 15599.

Del mismo modo se aplicó el modelo, a los datos para las MEDIANAS EMPRESAS para el año 2020, mirar en el **Anexo 2**. Teniendo en cuenta los datos arrojados por el software, observamos que en la siguiente tabla tenemos:

Cuadro 20: Información del modelo para los ingresos de las MEDIANAS EMPRESAS año 2020

Coefficients:				
	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	2.565e-10	6.323e-10	0.406	0.685
TRM_2020	1.909e-13	1.718e-13	1.111	0.267

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Como se observa en el Cuadro [20](#) obtenemos en relación de los ingresos 2020 con respecto a la TRM del mismo año, observamos una significancia 0,685 diferente de 0, por lo cual concluyo que es no es muy significativo los ingresos del presente año de las MEDIANAS EMPRESAS en Colombia ya que tenemos una confiabilidad del 31,5 % y el AIC: 15935.

Luego de dicha segmentación iniciamos analizando los ingresos asociados a las **MACROEMPRESAS** los cuales se presentan en el Cuadro [21](#).

Cuadro 21: Descriptiva de los ingresos de las MACROEMPRESAS de Colombia

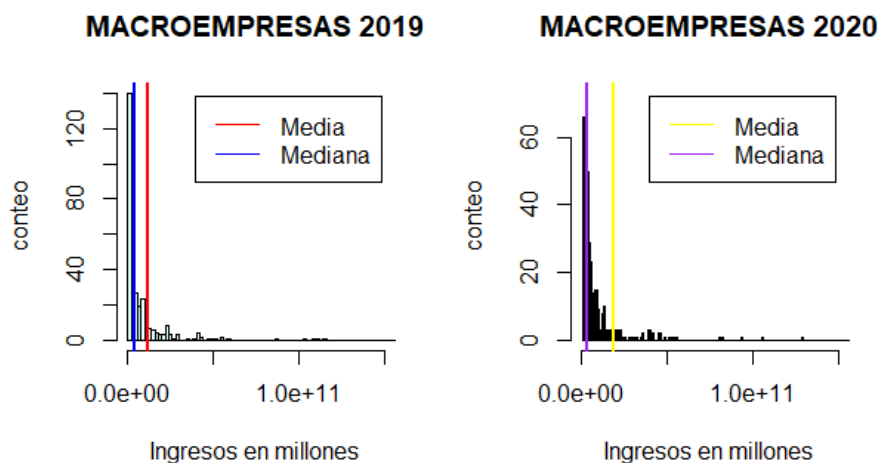
Variable	Ingresos 2019	Ingresos 2020
Media	18716211130	11482833154
Mediana	2973572809	3816516539
Varianza	5.049276e+21	1.113596e+21
Des. Estándar	71058260945	33370585977
IQR	8166677625	6287629466
Asimétrica	7.597395	9.504514
Curtosis	74.03492	117.0494

Fuente: Base de datos tomados de la página web datos abiertos Colombia

De acuerdo el Cuadro 21 Para el estudio realizado de las macroempresas se observa que los ingresos generados en el periodo 2019 presentaron ganancias en promedio de 18716211130 con una dispersión de 71058260945 mientras que en el 2020 se generó en promedio 11482833154 y con una dispersión de 33370585977. es decir, que las macroempresas en el año 2020 presentaron una pérdida de 30 puntos porcentuales con respecto a los ingresos del año 2019. También se observa que la mitad de las macroempresas empresas en el 2019 ganan más de 2973572809 mientras que en el 2020 tan solo la mitad presentaron ganancias mayores de 3816516539, esto presenta un balance crítico para el año 2020 (durante el covid-19). También se observa que ambas presentan una marcada asimetría positiva y leptocúrtica, es decir se logra evidenciar que para ambos periodos presentan datos agrupado que tienden a ingresos bajos. Ahora, en el periodo 2020 presenta menor dispersión de ingresos que el 2019 esto solo refleja que el 50 % de los ingresos están en 8166677625 mientras que en el 2020 es de 6287629466 Esto quiere decir que la llegada del covid-19 perjudico significativamente los ingresos de las macroempresas en Colombia.

A continuación, presentamos las correspondientes distribuciones de las variables de estudio, esto es, ingresos de las empresas para el año 2019 y 2020

Figura 16: Histograma de los ingresos de las MACROEMPRESAS de Colombia 2019 - 2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

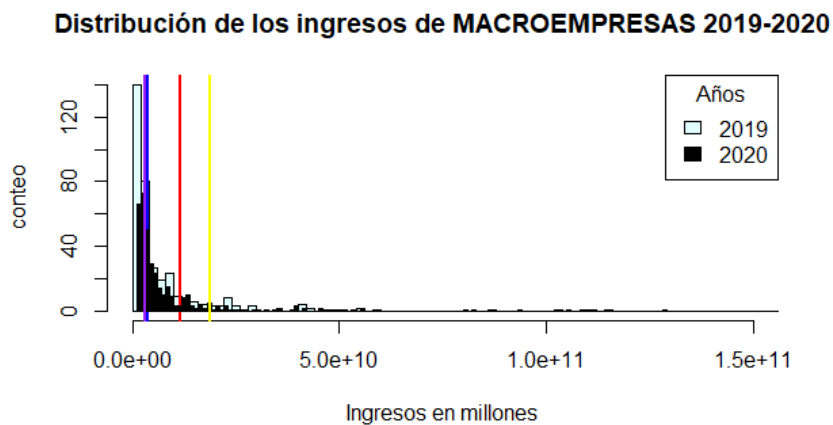
En la gráfica [16](#), se observa que la gráfica izquierda el comportamiento de los datos están muy relacionados a los presentados en el Cuadro [21](#), en efecto se evidencia una marcada asimetría positiva en los datos, también podemos observar que la línea roja representa la media para los datos de los ingresos del 2019 que es de 18716211130 de lo cual podemos decir que la mayoría de los ingresos de las microempresas están por debajo de la media, a diferencia de la línea azul representa la mediana para los datos de los ingresos del 2019 es decir que la mitad de las microempresas generan ingresos menores a 2973572809.

De igual manera, en la gráfica derecha se observa el comportamiento de los datos es heterogéneo, porque algunas empresas generan ingresos bajos y el resto de las empresas generan ingresos altos, también se observa que la línea amarilla representa la media para los datos de los ingresos del 2020 es decir que las empresas generan en promedio ingresos de 11482833154, a diferencia de la línea morada representa la mediana para los datos de los ingresos del 2020 es decir que la mitad de las microempresas generan ingresos menores a 3816516539.

A continuación, vamos a mirar la gráfica de los ingresos de las MACROEMPRESAS 2019 – 2020. En conclusión, al realizar el análisis de los indicadores de tendencia central es claro la robustez de la mediana frente a presencia de datos extremos o asimétricos



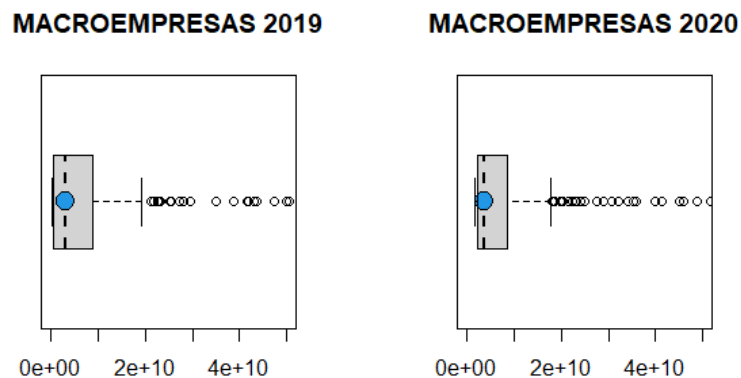
Figura 17: Comparación de los histogramas de los ingresos de las MACROEMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

En la gráfica [17](#), se observa lo siguiente hay una disminución en los ingresos de las empresas para el año 2020, eso se puede deducir ya que para ese año 2020 muchas empresas bajaron sus ingresos, por las medidas tomadas del gobierno COLOMBIANO. A diferencia del año 2019, teniendo en cuenta que hay una diferencia entre las medidas de tendencia central, la de color rojo y amarillo se hace referencia a la media respectivamente con los años comparados de igual forma el color azul y morado se hace referencia a la mediana respectivamente con los años comparados. A continuación, vamos a ver la siguiente grafica

Figura 18: Boxplot de los ingresos de las MACROEMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

En la figura 18 se observa una comparación de boxplot de macroempresas entre el año 2019 y 2020, lo cual se observa que en el año 2019 y 2020 los ingresos registrados por las empresas son asimétricos positivos muy sesgada a la izquierda con algunos ingresos mayores al rango máximo. Es decir, que algunas macroempresas generaron grandes ingresos en el año 2020 y 2019

De acuerdo con la metodología presentada para estimar los parámetros de la distribución gamma a través de software estadístico Rstudio encontramos que:

Cuadro 22: Estimaciones de los parámetros con respecto el año 2019 enfoque clásico

	alfa	beta
Momentos	1	5.342962e-11
MV constrOptim	1	5.342962e-11

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Cuadro 23: stimaciones de los parámetros con respecto el año 2020 enfoque clásico

	alfa	beta
Momentos	1	8.708652e-11
MV constrOptim	1	8.708652e-11

Fuente: Base de datos tomados de la página web datos abiertos Colombia

En el Cuadro [22](#) y [23](#) se observa el comportamiento de los datos con respecto  $\alpha$  y  $\beta$  con relación al enfoque clasico, mirar en el **Anexo 1**.

Cuadro 24: Estimación de los parámetros enfoque bayesiano

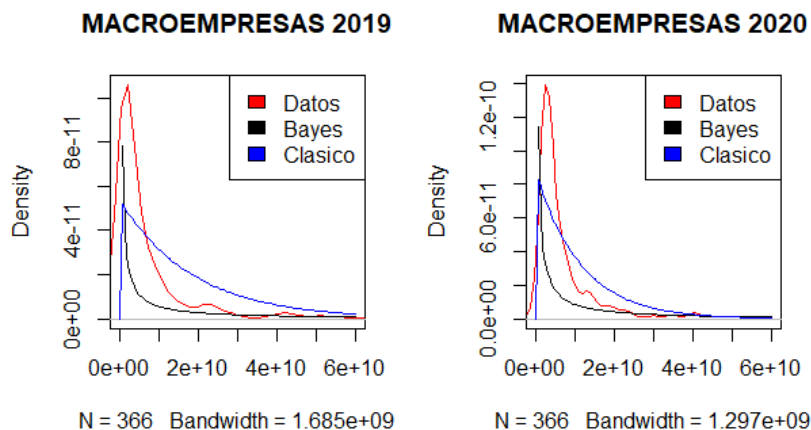
	Ingresos 2019	Ingresos 2020
alfa	0.0693756	0.1184051
beta	269780908829	96979203085

Fuente: Base de datos tomados de la página web datos abiertos Colombia

En el Cuadro [24](#) se observa el comportamiento de los datos con respecto  $\alpha$  y  $\beta$  con relación al enfoque bayesiano, mirar en el **Anexo 1**.

Teniendo en cuenta los datos arrojados por el software, utilizaremos la función gamma para ver su comportamiento con respecto a la densidad

Figura 19: Gamma (Enfoque Clásico y Enfoque Bayesiano) VS Densidad de los ingresos de las MACROEMPRESAS colombiana 2019–2020



Fuente: Base de datos tomados de la página web datos abiertos Colombia

De la figura [19](#) se observa dos graficas de las macroempresas colombianas del año 2019 y 2020, con enfoques clásico y bayesiano a través de la distribución gamma. Donde se observa que los ingresos no presentan mucha diferencia. En el año 2019 y 2020 presenta una curva de ingresos bajos que crece hacia ingresos altos y va decreciendo hasta llegar a pocas empresas que generan ingresos muy altos. además, se tiene que el comportamiento de la Gamma en los ingresos de las macroempresas en ambos años son muy diferentes con ambos enfoques.

Con el enfoque bayesiano se tiene una curva muy por debajo a la densidad de los ingresos de las macroempresas colombianas y con el enfoque clásico se tiene una curva decreciente por encima de la curva de los ingresos. Por lo tanto la que mejor se ajusta es el enfoque clásico tanto en el año 2019 como en el 2020.

Por eso se aplicó un modelo lineal generalizado GLM con respuesta Gamma, a los datos para las MACROEMPRESAS para el año 2019, mirar en el **Anexo 2**. Teniendo en cuenta los datos arrojados por el software, observamos que en la siguiente tabla tenemos:

Cuadro 25: Información del modelo para los ingresos de las MACROEMPRESAS año 2019

Coefficients:				
	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	5.295e-10	3.166e-10	1.672	0.0953
TRM_2019	-1.412e-13	9.441e-14	-1.496	0.1356

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Como se observa en el Cuadro [25](#) obtenemos en relación de los ingresos 2019 con respecto a la TRM del mismo año, observamos una significancia 0,0953 diferente de 0, por lo cual concluyo que es no es muy significativo los ingresos del presente año de las MACROEMPRESAS en Colombia ya que tenemos una confiabilidad del 90,47% y el AIC: 17596.

Del mismo modo se aplicó el modelo, a los datos para las MACROEMPRESAS para el año 2020, mirar en el **Anexo 2**. Teniendo en cuenta los datos arrojados por el software, observamos que en la siguiente tabla tenemos:

Cuadro 26: Información del modelo para los ingresos de las MACROEMPRESAS año 2020

Coefficients:				
	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	2.456e-10	2.414e-10	1.018	0.310
TRM_2020	-4.268e-14	6.450e-14	-0.662	0.509

Fuente: Base de datos tomados de la página web datos abiertos Colombia

Como se observa en el Cuadro [26](#) obtenemos en relación de los ingresos 2020 con respecto a la TRM del mismo año, observamos una significancia 0,310 diferente de 0, por lo cual concluyo que es no es muy significativo los ingresos del presente año de las MACROEMPRESAS en Colombia ya que tenemos una confiabilidad del 69% y el AIC: 17672.

# Conclusiones

- De acuerdo con los análisis realizados a las empresas colombianas en el año 2019 y 2020 se puede concluir que los ingresos reportados de las empresas en esos dos años varían de acuerdo con su tamaño. Si son micro, pequeña, mediana o macroempresas. Esto significa que la llegada del covid-19 afecto de forma diferente a esos grupos. En el caso de la micro y pequeña empresa en el año 2020 se observa que presento un gran déficit de ingresos salarial en comparación al año 2019, esto implica que afecto drásticamente a las personas menos asalariados, mientras que en la macro y mediana empresa a pesar de su disminución de ingresos en el 2020 no presento gran riesgo durante el covid-19.
- Al realizar las comparaciones del enfoque bayesiano y clásico. se obtuvo que ambas presentaron un comportamiento similar ante los ingresos de la micro, pequeña, mediana y macroempresas. Se observo que ambas generan una curva muy concentrada asimétrica a la derecha donde se obtiene que la mayoría de las empresas colombianas, generan ingresos bajos y muy pocos generan ingresos muy altos. Pero teniendo en cuenta el ajuste de la curva de los dos enfoques con respecto al grupo de empresas. Se obtiene que el enfoque bayesiano genera una visión más detallada y clara de los ingresos de las micro, pequeña, mediana y macroempresas de Colombia.
- Se utilizó el enfoque bayesiano y aplicando el modelo lineal generalizado con respuesta gamma se tuvieron cifras preocupantes, porque se obtuvo que todo el grupo de las empresas colombianas se comportó como la tasa representativa del mercado (TRM) en el año 2019. Pero en el año 2020 se comportó muy diferente. es decir, que hubo un decaimiento intensivo en los ingresos de las micro, pequeña, mediana y macroempresas de Colombia. Por lo tanto, la llegada del covid-19 en el año 2020 en Colombia, tuvo un impacto significativo para los ingresos de las empresas y especialmente a las micro y pequeñas empresas.

## Recomendación

- En un estudio previo, se recomienda analizar los ingresos de las mismas empresas en el año 2021 y 2020, para observar si después del 2020 (durante el covid-19) se presentó una disminución de empresas en Colombia.

# Anexo 1

## MICROEMPRESAS

```
> media<-mean(Base_de_datos$ 'ACTIVOS 2019')
> varianza<-var(Base_de_datos$ 'ACTIVOS 2019')
> alfa<-media^2/varianza
> alfa
[1] 1.420732
> beta<-varianza/media
> beta
[1] 12735435
```

Repetimos el mismo proceso para el año 2020

```
> media<-mean(Base_de_datos$ 'ACTIVOS 2020')
> varianza<-var(Base_de_datos$ 'ACTIVOS 2020')
> alfa<-media^2/varianza
> alfa
[1] 0.6710365
> beta<-varianza/media
> beta
[1] 13354120
```

Como valores iniciales podemos tomar los estimadores de momentos, que son:

$$\hat{\alpha} = \frac{\overline{X_n^2}}{S_n^2} \quad \text{y} \quad \hat{\beta} = \frac{\overline{X_n}}{S_n^2}$$

Primero calculamos los estadísticos  $\overline{X_n}$  y  $S_n^2$

```
> xb<-mean(Base_de_datos$ 'ACTIVOS 2019')
> s2<-mean((Base_de_datos$ 'ACTIVOS 2019'
+          -mean(Base_de_datos$ 'ACTIVOS 2019'))^2)
> xb
```



```
[1] 18093636
> s2
[1] 2.298007e+14
```

Ahora calculamos los estimadores de momentos

```
> al_mom <- xb^2 / s2
> la_mom <- xb / s2
```

Definimos la log-verosimilitud

```
> log.vero2 <- function(theta) - sum(log(dgamma(
Base_de_datos$'2019', theta[1], theta[2])))
```

Maximizamos con la función nlm

```
> Lmax2.1 <- nlm(log.vero2, c(al_mom, la_mom))
> Lmax2.1
```

\$minimum

```
[1] 6476.158
```

\$estimate

```
[1] 1.380963e+00 7.873620e-08
```

\$gradient

```
[1] -2.196413e-06 0.000000e+00
```

\$code

```
[1] 1
```

\$iterations

```
[1] 4
```

Maximizamos con la función constrOptim

```
> Lmax2.2 <- constrOptim(theta = c(al_mom, la_mom), f = log.vero2,
+                          grad = NULL, ui = diag(1, 2), ci = c(0, 0))
> Lmax2.2
```

\$par

```
[1] 1.290966e+00 7.137779e-08
```

```

$value
[1] 6475.383

$counts
function gradient
      163      NA

$convergence
[1] 0

$message
NULL

$outer.iterations
[1] 2

$barrier.value
[1] 9.612684e-05

```

De igual forma se repite para el año 2020.

Como valores iniciales podemos tomar los estimadores de momentos, que son:

$$\hat{\alpha} = \frac{\overline{X_n^2}}{S_n^2} \quad \text{y} \quad \hat{\beta} = \frac{\overline{X_n}}{S_n^2}$$

Primero calculamos los estadísticos  $\overline{X_n}$  y  $S_n^2$

```

> xb<-mean(Base_de_datos$ 'ACTIVOS 2020')
> s2<-mean((Base_de_datos$ 'ACTIVOS 2020'
+          -mean(Base_de_datos$ 'ACTIVOS 2020'))^2)
> xb
[1] 8961102
> s2
[1] 1.193407e+14

```

Ahora calculamos los estimadores de momentos

```

> al_mom <- xb^2 / s2
> la_mom <- xb / s2

```

Definimos la log-verosimilitud

```

> log.vero2 <- function(theta) - sum(log(dgamma(
Base_de_datos$ '2020', theta[1], theta[2])))

```

Maximizamos con la función nlm

```
> Lmax2.1 <- nlm(log.vero2, c(al_mom, la_mom))  
> Lmax2.1
```

```
$minimum  
[1] 6225.643
```

```
$estimate  
[1] 7.510452e-01 7.508841e-08
```

```
$gradient  
[1] 0.000121272 0.000000000
```

```
$code  
[1] 1
```

```
$iterations  
[1] 4
```

Maximizamos con la función constrOptim

```
> Lmax2.2 <- constrOptim(theta = c(al_mom, la_mom), f = log.vero2,  
+ grad = NULL, ui = diag(1, 2), ci = c(0, 0))  
> Lmax2.2
```

```
$par  
[1] 8.568154e-01 9.565275e-08
```

```
$value  
[1] 6222.026
```

```
$counts  
function gradient  
      167      NA
```

```
$convergence  
[1] 0
```

```
$message  
NULL
```

```
$outer.iterations
[1] 2
```

```
$barrier.value
[1] 9.892231e-05
```

## PEQUEÑAS EMPRESAS

```
> media<-mean(Base_de_datos$ 'ACTIVOS 2019')
> varianza<-var(Base_de_datos$ 'ACTIVOS 2019')
> alfa<-media^2/varianza
> alfa
[1] 2.826175
> beta<-varianza/media
> beta
[1] 64341020
```

Repetimos el mismo proceso para el año 2020

```
> media<-mean(Base_de_datos$ 'ACTIVOS 2020')
> varianza<-var(Base_de_datos$ 'ACTIVOS 2020')
> alfa<-media^2/varianza
> alfa
[1] 1.421418
> beta<-varianza/media
> beta
[1] 82882457
```

Como valores iniciales podemos tomar los estimadores de momentos, que son:

$$\hat{\alpha} = \frac{\overline{X_n^2}}{S_n^2} \quad \text{y} \quad \hat{\beta} = \frac{\overline{X_n}}{S_n^2}$$

Primero calculamos los estadísticos  $\overline{X_n}$  y  $S_n^2$

```
> xb<-mean(Base_de_datos$ 'ACTIVOS 2019')
> s2<-mean((Base_de_datos$ 'ACTIVOS 2019'
+
- mean(Base_de_datos$ 'ACTIVOS 2019'))^2)
> xb
[1] 181838984
> s2
[1] 1.166774e+16
```

Ahora calculamos los estimadores de momentos

```
> al_mom <- xb^2 / s2
> la_mom <- xb / s2
```

Definimos la log-verosimilitud

```
> log.vero2 <- function(theta) - sum(log(dgamma(
  Base_de_datos$'2019', theta[1], theta[2])))
```

Maximizamos con la función nlm

```
> Lmax2.1 <- nlm(log.vero2, c(al_mom, la_mom))
> Lmax2.1
```

\$minimum

```
[1] 7243.543
```

\$estimate

```
[1] 2.830429e+00 1.558477e-08
```

\$gradient

```
[1] 0.0002150002 0.0000000000
```

\$code

```
[1] 1
```

\$iterations

```
[1] 2
```

Maximizamos con la función constrOptim

```
> Lmax2.2 <- constrOptim(theta = c(al_mom, la_mom), f = log.vero2,
+                          grad = NULL, ui = diag(1, 2), ci = c(0, 0))
> Lmax2.2
```

\$par

```
[1] 2.833918e+00 1.558477e-08
```

\$value

```
[1] 7243.544
```

\$counts

```
[1] 0
```

```
$convergence
```

```
[1] 0
```

```
$message
```

```
NULL
```

```
$outer.iterations
```

```
[1] 1
```

```
$barrier.value
```

```
[1] -1.180613e-05
```

De igual forma se repite para el año 2020.

Como valores iniciales podemos tomar los estimadores de momentos, que son:

$$\hat{\alpha} = \frac{\overline{X_n^2}}{S_n^2} \quad \text{y} \quad \hat{\beta} = \frac{\overline{X_n}}{S_n^2}$$

Primero calculamos los estadísticos  $\overline{X_n}$  y  $S_n^2$

```
> xb<-mean(Base_de_datos$ 'ACTIVOS 2020 ')  
> s2<-mean((Base_de_datos$ 'ACTIVOS 2020 '  
+          -mean(Base_de_datos$ 'ACTIVOS 2020 '))^2)  
> xb  
[1] 117810633  
> s2  
[1] 9.737756e+15
```

Ahora calculamos los estimadores de momentos

```
> al_mom <- xb^2 / s2  
> la_mom <- xb / s2
```

Definimos la log-verosimilitud

```
> log.vero2 <- function(theta) - sum(log(dgamma(  
Base_de_datos$ '2020 ', theta[1], theta[2])))
```

Maximizamos con la función nlm

```
> Lmax2.1 <- nlm(log.vero2, c(al_mom, la_mom))  
> Lmax2.1
```

\$minimum

[1] 7171.294

\$estimate

[1] 1.312730e+00 1.209833e-08

\$gradient

[1] -0.0004257527 0.0000000000

\$code

[1] 1

\$iterations

[1] 4

Maximizamos con la función constrOptim

```
> Lmax2.2 <- constrOptim(theta = c(al_mom, la_mom), f = log.vero2,  
+                          grad = NULL, ui = diag(1, 2), ci = c(0, 0))  
> Lmax2.2
```

\$par

[1] 1.425313e+00 1.209833e-08

\$value

[1] 7173.793

\$counts

[1] 0

\$convergence

[1] 0

\$message

NULL

\$outer.iterations

[1] 1

\$barrier.value

[1] 9.201947e-05

## MEDIANAS EMPRESAS

```
> media<-mean(Base_de_datos$ 'ACTIVOS 2019')
> varianza<-var(Base_de_datos$ 'ACTIVOS 2019')
> alfa<-media^2/varianza
> alfa
[1] 2.734866
> beta<-varianza/media
> beta
[1] 501463328
```

Repetimos el mismo proceso para el año 2020

```
> media<-mean(Base_de_datos$ 'ACTIVOS 2020')
> varianza<-var(Base_de_datos$ 'ACTIVOS 2020')
> alfa<-media^2/varianza
> alfa
[1] 1.777952
> beta<-varianza/media
> beta
[1] 586061556
```

Como valores iniciales podemos tomar los estimadores de momentos, que son:

$$\hat{\alpha} = \frac{\overline{X_n^2}}{S_n^2} \quad \text{y} \quad \hat{\beta} = \frac{\overline{X_n}}{S_n^2}$$

Primero calculamos los estadísticos  $\overline{X_n}$  y  $S_n^2$

```
> xb<-mean(Base_de_datos$ 'ACTIVOS 2019')
> s2<-mean((Base_de_datos$ 'ACTIVOS 2019'
+          -mean(Base_de_datos$ 'ACTIVOS 2019'))^2)
> xb
[1] 1371435041
> s2
[1] 6.858454e+17
```

Ahora calculamos los estimadores de momentos

```
> al_mom <- xb^2 / s2
> la_mom <- xb / s2
```

Definimos la log-verosimilitud



```
> log.vero2 <- function(theta) - sum(log(dgamma(
Base_de_datos$ '2019', theta[1], theta[2])))
```

Maximizamos con la función constrOptim

```
> Lmax2.2 <- constrOptim(theta = c(al_mom, la_mom), f = log.vero2,
+ grad = NULL, ui = diag(1, 2), ci = c(0, 0))
> Lmax2.2
```

```
$par
[1] 2.818607e+00 2.054883e-09
```

```
$value
[1] 7977.694
```

```
$counts
function gradient
      252      NA
```

```
$convergence
[1] 0
```

```
$message
NULL
```

```
$outer.iterations
[1] 3
```

```
$barrier.value
[1] -8.045022e-06
```

De igual forma se repite para el año 2020.

Como valores iniciales podemos tomar los estimadores de momentos, que son:

$$\hat{\alpha} = \frac{\overline{X_n^2}}{S_n^2} \quad \text{y} \quad \hat{\beta} = \frac{\overline{X_n}}{S_n^2}$$

Primero calculamos los estadísticos  $\overline{X_n}$  y  $S_n^2$

```
> xb<-mean(Base_de_datos$ 'ACTIVOS 2020')
> s2<-mean((Base_de_datos$ 'ACTIVOS 2020'
+ -mean(Base_de_datos$ 'ACTIVOS 2020'))^2)
```

```
> xb
[1] 1041989034
```

```
> s2
[1] 6.090012e+17
```

Ahora calculamos los estimadores de momentos

```
> al_mom <- xb^2 / s2
> la_mom <- xb / s2
```

Definimos la log-verosimilitud

```
> log.vero2 <- function(theta) - sum(log(dgamma(
Base_de_datos$'2020', theta[1], theta[2])))
```

Maximizamos con la función constrOptim

```
> Lmax2.2 <- constrOptim(theta = c(al_mom, la_mom), f = log.vero2,
+                          grad = NULL, ui = diag(1, 2), ci = c(0, 0))
> Lmax2.2
```

```
$par
[1] 2.808335e+00 2.047441e-09
```

```
$value
[1] 7977.724
```

```
$counts
function gradient
      1492      NA
```

```
$convergence
[1] 0
```

```
$message
NULL
```

```
$outer.iterations
[1] 13
```

```
$barrier.value
[1] -6.762422e-06
```

## MACROEMPRESAS

```
> media<-mean(Base_de_datos$ 'ACTIVOS 2019')
> varianza<-var(Base_de_datos$ 'ACTIVOS 2019')
> alfa<-media^2/varianza
> alfa
[1] 0.0693756
> beta<-varianza/media
> beta
[1] 269780908829
```

Repetimos el mismo proceso para el año 2020

```
> media<-mean(Base_de_datos$ 'ACTIVOS 2020')
> varianza<-var(Base_de_datos$ 'ACTIVOS 2020')
> alfa<-media^2/varianza
> alfa
[1] 0.1184051
> beta<-varianza/media
> beta
[1] 96979203085
```

## Anexo 2

### MICROEMPRESAS

```
> TRM_2019<-c(Datos_TRM$TRM2019)
> TRM_2019=TRM_2019[!is.na(TRM_2019)]
> ingresos2019<-c(Base_de_datos$'ACTIVOS 2019'[1:365])
> Datos<-data.frame(ingresos2019,TRM_2019)
> Modelo1<-glm(ingresos2019~TRM_2019, family = Gamma,data = Datos)
```

veremos el resumen del modelo

```
> summary(Modelo1)
```

Call:

```
glm(formula = ingresos2019 ~ TRM_2019, family = Gamma, data = Datos)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7461	-0.5496	-0.2903	0.2115	1.7975

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.033e-06	6.003e-08	17.21	<2e-16 ***
TRM_2019	-2.910e-10	1.746e-11	-16.67	<2e-16 ***

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

(Dispersion parameter for Gamma family taken to be 0.5415776)

Null deviance: 316.87 on 364 degrees of freedom  
Residual deviance: 169.83 on 363 degrees of freedom  
AIC: 12669

Number of Fisher Scoring iterations: 6

Del mismo modo aplicaremos el mismo proceso anterior para analizar los datos arrojado para los ingresos de las MICROEMPRESAS de Colombia en el periodo 2020.

```
> TRM_2020<-c(Datos_TRM$TRM2020)
> TRM_2020=TRM_2020[!is.na(TRM_2020)]
> ingresos2020<-c(Base_de_datos$ 'ACTIVOS 2020')
> Datos<-data.frame(ingresos2020 ,TRM_2020)
> Modelo1<-glm(ingresos2020~TRM_2020, family = Gamma,data = Datos)
veremos el resumen del modelo
> summary(Modelo1)
```

Call:

```
glm(formula = ingresos2020 ~ TRM_2020, family = Gamma, data = Datos)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6679	-1.3374	-0.7355	0.3160	2.2896

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.341e-08	1.207e-07	0.194	0.846
TRM_2020	2.394e-11	3.280e-11	0.730	0.466

(Dispersion parameter for Gamma family taken to be 1.467457)

Null deviance: 502.89 on 365 degrees of freedom  
Residual deviance: 502.12 on 364 degrees of freedom  
AIC: 12456

Number of Fisher Scoring iterations: 7

## PEQUEÑAS EMPRESAS

```
> TRM_2019<-c(Datos_TRM$TRM2019)
> TRM_2019=TRM_2019[!is.na(TRM_2019)]
> ingresos2019<-c(Base_de_datos$ 'ACTIVOS 2019'[1:365])
> Datos<-data.frame(ingresos2019 ,TRM_2019)
```

```

> Modelo1<-glm(ingresos2019-TRM_2019, family = Gamma, data = Datos)
veremos el resumen del modelo
> summary(Modelo1)

Call:
glm(formula = ingresos2019 ~ TRM_2019, family = Gamma, data = Datos)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8457  -0.3417  -0.1299   0.2054   1.2855

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.480e-08  3.697e-09   20.23  <2e-16 ***
TRM_2019    -2.076e-11  1.090e-12  -19.05  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.2032695)

Null deviance: 136.705  on 364  degrees of freedom
Residual deviance:  60.219  on 363  degrees of freedom
AIC: 14138

Number of Fisher Scoring iterations: 5

Del mismo modo aplicaremos el mismo proceso anterior para analizar los datos
arrojado para los ingresos de las PEQUEÑAS EMPRESAS de Colombia en el
periodo 2020.
> TRM_2020<-c(Datos_TRM$TRM2020)
> TRM_2020=TRM_2020[!is.na(TRM_2020)]
> ingresos2020<-c(Base_de_datos$'ACTIVOS 2020')
> Datos<-data.frame(ingresos2020,TRM_2020)
> Modelo1<-glm(ingresos2020-TRM_2020, family = Gamma, data = Datos)
veremos el resumen del modelo
> summary(Modelo1)

Call:
glm(formula = ingresos2020 ~ TRM_2020, family = Gamma, data = Datos)

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7504	-0.8071	-0.2912	0.4401	1.5192

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.029e-09	6.363e-09	0.790	0.430
TRM.2020	9.380e-13	1.725e-12	0.544	0.587

(Dispersion parameter for Gamma family taken to be 0.6936018)

Null deviance: 372.76 on 365 degrees of freedom  
Residual deviance: 372.56 on 364 degrees of freedom  
AIC: 14343

Number of Fisher Scoring iterations: 6

## MEDIANAS EMPRESAS

```
> TRM_2019<-c(Datos.TRM$TRM2019)
> TRM_2019=TRM_2019[!is.na(TRM_2019)]
> ingresos2019<-c(Base_de_datos$'ACTIVOS 2019'[1:365])
> Datos<-data.frame(ingresos2019,TRM_2019)
> Modelo1<-glm(ingresos2019~TRM_2019, family = Gamma,data = Datos)
veremos el resumen del modelo
> summary(Modelo1)
```

Call:

```
glm(formula = ingresos2019 ~ TRM_2019, family = Gamma, data = Datos)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.7602	-0.3176	-0.1509	0.1789	1.3895

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.736e-09	4.969e-10	19.59	<2e-16 ***
TRM.2019	-2.698e-12	1.465e-13	-18.42	<2e-16 ***

---

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

(Dispersion parameter for Gamma family taken to be 0.2091281)

Null deviance: 130.651 on 364 degrees of freedom  
Residual deviance: 57.072 on 363 degrees of freedom  
AIC: 15599

Number of Fisher Scoring iterations: 5

Del mismo modo aplicaremos el mismo proceso anterior para analizar los datos arrojado para los ingresos de las MEDIANAS EMPRESAS de Colombia en el periodo 2020.

```
> TRM_2020<-c(Datos_TRM$TRM2020)
> TRM_2020=TRM_2020[!is.na(TRM_2020)]
> ingresos2020<-c(Base_de_datos$'ACTIVOS 2020')
> Datos<-data.frame(ingresos2020,TRM_2020)
> Modelo1<-glm(ingresos2020~TRM_2020, family = Gamma,data = Datos)
```

veremos el resumen del modelo

```
> summary(Modelo1)
```

Call:

```
glm(formula = ingresos2020 ~ TRM_2020, family = Gamma, data = Datos)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.4584	-0.7272	-0.2103	0.4319	1.3013

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.565e-10	6.323e-10	0.406	0.685
TRM_2020	1.909e-13	1.718e-13	1.111	0.267

(Dispersion parameter for Gamma family taken to be 0.543106)

Null deviance: 356.15 on 365 degrees of freedom  
Residual deviance: 355.49 on 364 degrees of freedom  
AIC: 15935



Number of Fisher Scoring iterations: 6

## MACROEMPRESAS

```
> TRM_2019<-c(Datos_TRM$TRM2019)
> TRM_2019=TRM_2019[!is.na(TRM_2019)]
> ingresos2019<-c(Base_de_datos$'ACTIVOS 2019'[1:365])
> Datos<-data.frame(ingresos2019,TRM_2019)
> Modelo1<-glm(ingresos2019~TRM_2019, family = Gamma,data = Datos)
veremos el resumen del modelo
> summary(Modelo1)
```

```
Call:
glm(formula = ingresos2019 ~ TRM_2019, family = Gamma, data = Datos1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7551	-2.0419	-1.2260	-0.7875	7.5100

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.295e-10	3.166e-10	1.672	0.0953 .
TRM_2019	-1.412e-13	9.441e-14	-1.496	0.1356

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

(Dispersion parameter for Gamma family taken to be 12.99136)

Null deviance: 1209.9 on 364 degrees of freedom  
Residual deviance: 1180.2 on 363 degrees of freedom  
AIC: 17596

Number of Fisher Scoring iterations: 10

# Referencias

- [1] Griselda Dávila Aragón, Francisco Ortiz Arango, and Fernando Cruz Aranda. Cálculo del valor en riesgo operacional mediante redes bayesianas para una empresa financiera. *Contaduría y administración*, 61(1):176–201, 2016.
- [2] Liliana Blanco Castañeda. *Probabilidad*. Univ. Nacional de Colombia, 2004.
- [3] Federico Cingano. Trends in income inequality and its impact on economic growth. 2014.
- [4] Edinson Javier Díaz Prieto. Estimación de un modelo lineal generalizado mixto para datos de conteo con exceso de ceros. 2018.
- [5] Guadalupe Gómez and Pedro Delicado. Curso de inferencia y decisión. *Barcelona: Universidad Politécnica de Cataluña*, 2006.
- [6] Gustavo Hernández and Francisco Javier Lasso. Estimación de la relación entre salario mínimo y empleo en colombia: 1984-2000. *Revista de Economía del Rosario*, 6(2):117–138, 2003.
- [7] Chun-Chang Lee, Chih-Min Liang, Jian-Zheng Chen, and Cheng-Huang Tung. Effects of the housing price to income ratio on tenure choice in taiwan: forecasting performance of the hierarchical generalized linear model and traditional binary logistic regression model. *Journal of Housing and the Built Environment*, 33(4):675–694, 2018.
- [8] Peter McCullagh and John A Nelder. *Generalized linear models*. Routledge, 2019.
- [9] Mónica Parra. Tendencias recientes en la distribución del ingreso en colombia. 1999.
- [10] Christian P Robert et al. *The Bayesian choice: from decision-theoretic foundations to computational implementation*, volume 2. Springer, 2007.
- [11] Fabio Sánchez. Desigualdad del ingreso 1976-2004. 2005.
- [12] Jorge Edgar Silva Veloza. Estudio del índice de precios al consumidor en el departamento del tolima y departamentos aledaños mediante general algebraic modeling system. 2018.

- [13] Ferney Arturo Simbasica Numpaque. Medición de los ingresos en bogotá mediante modelos lineales generalizados con distribución tweedie para el año 2017. 2019.
- [14] Gloria Pascuala Chiquito Tigua, Ninive Victoria Plua Parrales, and Jessenia H Morán Chilán. Red bayesiana para analizar los mecanismos de adaptación de los comerciantes ante los retos de la covid-19. *3c Empresa: investigación y pensamiento crítico*, (1):67–81, 2020.
- [15] Camilo Ernesto Tovar, José Antonio Ocampo, and Fabio Sánchez Torres. Mercado laboral y distribución del ingreso en colombia en los años noventa. *Revista de la CEPAL*, 2000.