



**Acta de Correcciones al Proyecto de Grado**  
**Matemáticas Aplicadas**

**Fecha:** 25/03/2021

**Autores:** Natalia Vanegas Torres

**Nombre del Proyecto de Grado:**

Estimación de la curva y pronóstico del número de infectados por COVID-19 mediante modelos estadísticos en Colombia durante el año 2020

**Director:** Daniel Enrique González Gómez

Como indica el artículo 2.27 de las Directrices de Trabajo de Grado, he verificado que los estudiantes indicados arriba han implementado todas las correcciones que los Jurados del Proyecto de Grado definieron que se efectuaran, como consta en el Acta de Calificación correspondiente.

\_\_\_\_\_  
Firma de Director(a) del Proyecto de Grado

Nota de Aceptación

Aprobado por el Comité de Trabajo de Grado  
en cumplimiento de los requisitos exigidos por la  
Pontificia Universidad Javeriana para optar el  
título de Profesional en Matemáticas Aplicadas.



---

**Dr Camilo Rocha Niño**  
Decano de la Facultad de Ingeniería



---

**Dra. Diana H. Bueno Carreño**  
Director Carrera Matemáticas Aplicadas



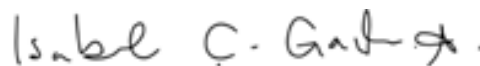
---

**Daniel Enrique González Gómez**  
Director(a) Trabajo



---

**David Arango Londoño**  
Jurado 1



---

**Isabel García Arboleda**  
Jurado 2

Señores

**Pontificia Universidad Javeriana Cali**

Dra. Diana Haidive Bueno Carreño

Directora Matemáticas Aplicadas Cali

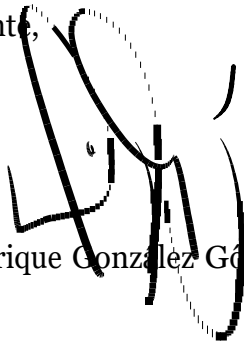
Febrero 9, 2021

Cordial saludo.

Por medio de la presente confirmo que he revisado el proyecto de grado titulado **Estimación de la curva y pronóstico del número de infectados por COVID-19 mediante modelos estadísticos en Colombia durante el año 2020**, escrito por la estudiante Natalia Vanegas Torres del programa de Matemáticas Aplicadas identificada con ID: 8935154 y que en calidad de director apruebo que sea entregado para su evaluación.

Gracias.

Atentamente,

A handwritten signature in black ink, appearing to read 'Daniel Enrique González Gómez', is written over a faint, dotted-line signature.

Daniel Enrique González Gómez

Señores

**Pontificia Universidad Javeriana Cali**

Dra. Diana Haidive Bueno Carreño

Directora Matemáticas Aplicadas Cali

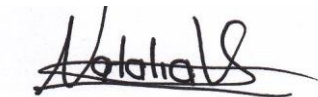
Febrero 9, 2021

Cordial saludo.

Me permito presentar a su consideración el proyecto de grado denominado “**Estimación de la curva y pronóstico del número de infectados por COVID-19 mediante modelos estadísticos en Colombia durante el año 2020**”, con el fin de cumplir con los requisitos exigidos por la Universidad para optar por el título de Profesional en Matemáticas Aplicadas.

Gracias.

Atentamente,



Natalia Vanegas Torres

C.C. 1113696814

ID: 8935154

**Estimación de la curva y pronóstico del número de infectados por  
COVID-19 mediante modelos estadísticos en Colombia durante el año  
2020.**

**NATALIA VANEGAS TORRES**

**PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERÍA Y CIENCIAS  
SANTIAGO DE CALI  
2021**

**NATALIA VANEGAS TORRES**

**Trabajo de grado presentado como requisito para optar por el título de  
profesional en Matemáticas Aplicadas**

**Director de trabajo de grado**

**DANIEL ENRIQUE GONZÁLEZ GÓMEZ**

**PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERÍA Y CIENCIAS  
SANTIAGO DE CALI**

**2021**

# Índice

<b>RESUMEN</b>	<b>5</b>
<b>ABSTRACT</b>	<b>6</b>
<b>1. INTRODUCCIÓN</b>	<b>7</b>
<b>2. DESCRIPCIÓN DEL PROBLEMA</b>	<b>9</b>
2.1. DEFINICIÓN DEL PROBLEMA . . . . .	9
2.2. OBJETIVOS . . . . .	9
2.2.1. OBJETIVO GENERAL . . . . .	9
2.2.2. OBJETIVOS ESPECÍFICOS . . . . .	9
2.3. ALCANCE . . . . .	10
2.4. JUSTIFICACIÓN . . . . .	10
<b>3. DESARROLLO DEL PROYECTO</b>	<b>11</b>
3.1. ANTECEDENTES . . . . .	11
3.2. MARCO TEÓRICO . . . . .	14
3.2.1. Tipos de pronóstico . . . . .	17
3.3. Métodos cuantitativos . . . . .	18
3.3.1. Métodos de series de tiempo . . . . .	18
3.4. RESULTADOS OBTENIDOS . . . . .	27
3.4.1. Presentación de variables . . . . .	27
3.4.2. Análisis univariado . . . . .	28
3.4.3. Análisis descriptivo bivariado . . . . .	33
3.5. Generación de Modelos . . . . .	38
<b>4. CONCLUSIONES</b>	<b>51</b>
<b>5. ANEXOS</b>	<b>52</b>
5.1. Código principal . . . . .	52
5.2. Modelo ARIMA . . . . .	57
5.3. Gráficos . . . . .	58
5.4. Dashboard . . . . .	64

# Índice de figuras

1. Dashboard de la Universidad Jhons Hopkins. Fecha de captura: 5 de Mayo de 2020 . . . . .	12
2. Dashboard de la Organización Mundial de la Salud. Fecha de captura: 5 de Mayo de 2020. Tomado de (16) . . . . .	13

3.	Dashboard de la HealthMap. Fecha de captura: 5 de Mayo de 2020.	
	Tomado de (17)	13
4.	Pasos a seguir en la elaboración de un pronóstico	15
5.	Pronósticos de acuerdo a su tendencia. Construcción propia.	18
6.	Componenetes de una serie de tiempo. Tomado de: Universidad Carlos	
	III (28)	19
7.	Tipo de atención. Tomado de: INS (25)	28
8.	Tipo de contagio. Tomado de: INS (25)	29
9.	Contagiados por rango de edad. Tomado de: INS (25)	30
10.	Cantidad de muertos por edad. Tomado de: INS (25)	33
11.	Cantidad de recuperados por edad. Tomado de: INS (25)	34
12.	Pirámide de contagios por edad y sexo. Tomado de: INS (25)	35
13.	Curva de contagiados en Colombia durante el año 2020	36
14.	Fechas destacadas durante la pandemia	37
15.	Pronóstico modelo Holt-Winters	39
16.	Pronóstico modelo Holt	40
17.	Pronóstico modelo Medias Móviles para $n = 3$	42
18.	Función de autocorrelación	43
19.	Función de autocorrelación parcial	44
20.	Pronósticos mediante el modelo ARIMA	45
21.	Propuesta de modelo con la función auto.arima del paquete <i>forecast</i> de R.	46
22.	Datos reales vs modelos	47
23.	Pronóstico realizado por SALUTIA. Tomado de: Organización Salutia (24)	48
24.	Dashboard de Colombia, 31 de diciembre de 2020. <a href="https://nativanegas.github.io/index.html">https://nativanegas.github.io/index.html</a>	
		50

## Índice de cuadros

1.	Variables. Tomado de: INS (25)	27
2.	Cantidad de contagiados por departamento. Tomado de: INS (25)	31
3.	Cantidad de fallecidos por departamento. Tomado de: INS (25)	32
4.	Pronóstico modelo Holt-Winters	39
5.	Pronóstico modelo Holt	40
6.	Errores de medias móviles	41
7.	Pronóstico de medias móviles	41
8.	Pronóstico mediante modelo ARIMA	45
9.	Errores de los modelos	46
10.	Comparación de pronósticos	47

## RESUMEN

Este trabajo de investigación presenta la importancia de los modelos estadísticos en la elaboración de pronósticos a corto plazo de enfermedades contagiosas como lo es el COVID-19 en Colombia durante el año 2020.

Para realizar el trabajo se descargaron los datos suministrados por el Instituto Nacional de Salud desde el portal de datos abiertos Colombia de manera sencilla y ordenada. Tomando la cantidad de confirmados diarios como una serie de tiempo se realizó un análisis de su tendencia y estacionalidad para luego realizar el pronóstico desde el 22 de diciembre de 2020 hasta el día 31 de diciembre de 2020. Los modelos seleccionados para este pronóstico fueron: Holt, Holt-Winters, Medias Móviles y ARIMA, luego se realizó una validación con los errores de: Desviación Absoluta de la Media, Error Cuadrático Medio y Raíz del Error Cuadrático Medio para determinar cual es el mejor modelo que pronostica la curva de contagiados diarios.

**Palabras claves:** coronavirus, modelos de pronóstico, modelo Holt-Winters, modelo Holt, modelo ARIMA, modelo de medias móviles.

## **ABSTRACT**

This research work presents the importance of statistical models in the preparation of short-term forecasts of contagious diseases such as COVID-19 in Colombia during the year 2020.

To carry out the work, the data provided by the National Institute of Health was downloaded from datos abiertos Colombia in a simple and orderly manner. Taking the number of confirmed daily as a time series, an analysis of its trend and seasonality was carried out and then the forecast was made from December 22, 2020 to December 31, 2020. The models selected for this forecast were: Holt, Holt-Winters, Moving Averages and ARIMA, then a validation was performed with the errors of: Absolute Deviation of the Mean, Mean Square Error and Root of Mean Square Error to determine which is the best model that predicts the daily contagion curve.

# 1. INTRODUCCIÓN

Según la OMS (10) en diciembre del año 2019 fue detectado por primera vez en el mundo el virus Sars-CoV-2, el cual es un tipo de coronavirus causante de la enfermedad COVID-19 en la ciudad de Wuhan, China.

Los coronavirus son una extensa familia de virus, los cuales causan infecciones respiratorias que pueden ir desde un resfriado común hasta enfermedades más graves como el síndrome respiratorio de Oriente Medio (MERS) y el síndrome respiratorio agudo severo (SRAS).

El COVID-19 ha tenido una propagación significativa a nivel global llegando a gran parte de los países del mundo, por lo cual, el 11 de marzo del 2020 la Organización Mundial de la Salud la declaró pandemia (16).

Esta enfermedad se propaga de persona a persona a través de las gotículas procedentes de la nariz o la boca que salen despedidas cuando una persona infectada tose o exhala. También debido a que esas gotículas quedan en objetos o superficies al tener contacto con ellas y posteriormente tener contacto con los ojos, la nariz o la boca las personas se pueden contagiar.

Los síntomas más comunes del COVID-19 son fiebre, cansancio y tos seca. Algunas personas pueden presentar dificultad para respirar; aunque la mayoría de las personas, según la OMS alrededor del 80 %, se recupera de la enfermedad sin necesidad de realizar ningún tratamiento especial.

Es importante resaltar que las personas mayores y las que padecen afecciones médicas subyacentes, como hipertensión arterial, problemas cardiacos o diabetes, tienen más probabilidades de desarrollar una enfermedad grave, lo cual ha sido la causa de muertes en muchos de los casos positivos.

Aunque la letalidad del virus comparada con otros como el SRAS y el MERS ha sido menor, el nivel de propagación del Sars-Cov-2 ha sido el mayor motivo de preocupación, puesto que desde diciembre del año 2019 al 31 de diciembre de 2020 el número de contagios y muertes en el mundo según (9) era de 81.485.119 y 1.798.754 respectivamente. En Colombia, el primer caso de contagio lo confirmó el Ministerio de Salud y Protección Social el 6 de marzo de 2020 en Bogotá, desde entonces el virus empezó a propagarse de manera significativa por lo cual, el día 24 de marzo el presidente Iván Duque declaró aislamiento social con el objetivo de disminuir el número de contagios y muertes, el aislamiento se extendió hasta el día 30 de agosto de 2020 fecha en la cual se dio inicio al aislamiento inteligente que consistía en mantener las restricciones a los adultos mayores y al sistema educativo junto al distanciamiento social y las normas de bioseguridad.

Durante el periodo de cuarentena estricta se presentó el pico de contagios impulsado principalmente por el día sin IVA propuesto por el presidente Iván Duque el 19 de junio. La cantidad de contagios diarios para el día 31 de diciembre de 2020 en Colombia había un total de 1.642.775 contagiados y 43.213 muertos según los datos suministrados por el Instituto Nacional de Salud de Colombia (I11).

El presente trabajo de grado está dividido en 4 partes: Definición del problema donde se plantea la necesidad de utilizar modelos estadísticos con el fin de conocer el crecimiento del COVID-19 a nivel nacional basado en los datos recolectados del Instituto Nacional de Salud. En el alcance se presenta de manera resumida los objetivos del trabajo de grado, especificando qué se espera pronosticar y evidenciar de manera gráfica mediante un dashboard la cantidad de contagiados, muertos y recuperados a través de modelos estadísticos.

En la justificación se da a conocer la importancia de realizar este trabajo de investigación el cual tiene como fin estudiar el comportamiento de esta pandemia durante el año 2020 y de esta manera saber las zonas con mayor propagación de la enfermedad en Colombia. En la sección de antecedentes se presentan los dashboards que han elaborado hasta el momento diversas entidades reconocidas a nivel mundial enunciados en (I15), (I16), (I17) los cuales serán usados de referencia para la elaboración del dashboard a nivel nacional, además de presentar estudios previos acerca de pronósticos estadístico-matemáticos sobre el COVID-19.

En el marco teórico se presentan la información base con la cual se trabaja en el proyecto de investigación, en este se definen los métodos de pronósticos, los pasos a seguir y se describen algunos de los modelos más importantes. En los resultados obtenidos se presentan los pronósticos realizados con los modelos *Holt-Winters*, *Holt*, *ARIMA* y *Medias Móviles* los cuales fueron seleccionados para compararlos con los datos reales y saber mediante la medición de los errores de Desviación Absoluta de la Media, Error Medio Cuadrático y Raíz Cuadrática del Error Medio cuál modelo aproxima mejor los datos finales al 31 de diciembre de 2020. Cabe destacar que los datos tomados son una muestra de los contagiados totales dado que muchas personas contagiadas son asintomáticas al virus y otras a pesar de estar enfermas con síntomas no recurren a los centros de salud para realizarse la respectiva prueba, por lo que los datos tomados son solo de la población que se realizó las pruebas que hace el gobierno.

## **2. DESCRIPCIÓN DEL PROBLEMA**

### **2.1. DEFINICIÓN DEL PROBLEMA**

Debido a la rapidez de propagación del Sars-Cov-2 a nivel mundial las entidades gubernamentales han tomado decisiones importantes de la manera más rápida posible con el fin de hacer frente a esta pandemia y así evitar que en poco tiempo gran parte de la población se contagie y colapse la red hospitalaria.

En diferentes estudios se han utilizado modelos basados principalmente en ecuaciones diferenciales estocásticas. En el presente trabajo de investigación se tratará el problema mediante métodos de series de tiempo para pronosticar en cada instante de tiempo la cantidad de contagios y de esta manera brindar a la población una herramienta para que puedan conocer la situación real del país frente a la cantidad de contagiados que se presentan.

### **2.2. OBJETIVOS**

#### **2.2.1. OBJETIVO GENERAL**

Estimar y pronosticar por medio de métodos estadísticos la curva de propagación del COVID-19 durante el año 2020 en Colombia eligiendo el mejor método que permita pronosticar el número de contagios diarios con el fin de brindar una herramienta informativa a los colombianos.

#### **2.2.2. OBJETIVOS ESPECÍFICOS**

- Realizar una revisión de las principales metodologías para una generación de pronósticos con el fin de ser utilizados en la construcción de modelos de la serie de tiempo generada por la transmisión del virus COVID-19.
- Construir e identificar diferentes modelos de pronóstico con los datos obtenidos a nivel nacional para analizar el comportamiento del COVID-19 en Colombia.
- Realizar un análisis comparativo de los modelos de pronóstico seleccionados para escoger el que mejor se ajuste a la serie a partir de indicadores estadísticos y representarlo gráficamente.
- Generar una estrategia de comunicación que permita a los ciudadanos visualizar la cantidad de contagiados, muertos y recuperados por COVID-19 en los departamentos de Colombia durante el año 2020.

### **2.3. ALCANCE**

El presente trabajo está dirigido a la construcción y comparación de modelos a través de diferentes técnicas de pronóstico estadísticas que permitan la visualización del comportamiento de la evolución del COVID-19 a corto plazo, con el fin de realizar varias modelaciones y escoger el modelo de pronóstico óptimo que permita de manera eficiente evidenciar el comportamiento del COVID-19 en Colombia a lo largo del año 2020 realizando su representación gráfica a través de un dashboard para poder contribuir a estudios sobre el comportamiento del mismo.

### **2.4. JUSTIFICACIÓN**

Conocer el comportamiento y las principales características de la curva de propagación de una enfermedad como el COVID-19 es de gran importancia para las entidades gubernamentales y la población dado que esto permite tomar medidas tanto preventivas como correctivas para contrarrestar la propagación de la enfermedad. Es por ello que la estadística se vuelve una alternativa para el manejo de la información en la toma de decisiones, dado que desde esta área de las matemáticas se pueden crear modelos de pronósticos que permitan conocer el alcance de la enfermedad y la influencia de diferentes factores en la propagación de la misma.

De esta manera, el presente trabajo de investigación pretende construir y describir un modelo de predicción estadística para la determinación de la cantidad de contagiados que tendrá el COVID-19 al día 31 de diciembre de 2020, el comportamiento de la evolución del virus en cada departamento del país a lo largo del año y presentando de manera visual los datos en un dashboard en tiempo real para una mayor comprensión de los mismos, con el objetivo de verificar la efectividad de las medidas gubernamentales tomadas frente a la epidemia en Colombia y tener un referente de estudio en caso de un nuevo virus.

## 3. DESARROLLO DEL PROYECTO

### 3.1. ANTECEDENTES

En el momento que se conoció a nivel mundial de la nueva pandemia se empezaron a realizar diferentes estudios de sus posibles causas y se presentaron modelos de proyección de su propagación, entre los cuales se destacan los siguientes:

En Pérez et al. (21) se presentan características clínico-epidemiológicas del COVID-19, en este trabajo los autores describieron los síntomas de la enfermedad. Siendo los síntomas más comunes: fiebre, tos, disnea, mialgia y fatiga, también se presenta pérdida de olfato y gusto aunque en los casos graves se presenta neumonía, síndrome de dificultad respiratoria aguda y choque séptico. El 30 de enero de 2020 la OMS la declaró como una emergencia sanitaria de preocupación internacional y el 11 de marzo la reconoció como una pandemia.

Hasta la fecha se conocen seis especies de coronavirus, cuatro de ellas causan síntomas comunes de gripa en personas inmunodeprimidas (229E, OC43, NL63 y HKU1) y dos especies que causan síndrome respiratorio agudo severo con altas tasas de mortalidad.

El mecanismo de transmisión animal-humano, aunque es desconocido se observa que se puede dar por las secreciones respiratorias o materias procedente del aparato digestivo. En el caso de la transmisión humano-humano, se observa que se produce por contacto con gotas respiratorias de más de 5 micras y las manos u objetos contaminadas por dichas secreciones. Este virus está asociado a una alta morbimortalidad en pacientes de la tercera edad o con enfermedades crónicas preexistentes, en la mayoría de los casos se presenta un cuadro clínico correspondiente a una infección respiratoria aguda.

En la Universidad del Norte de Barranquilla Yanes y Porras (22), proponen diferentes modelos estadísticos y matemáticos que ayudan a tomar decisiones de política pública frente a esta enfermedad. El primer modelo que utilizan es el modelo SIR (Susceptibles, Infectados, Recuperados), este modelo ha sido el más importante en los últimos años ya que trata de proyectar cuál será el número de personas infectadas en una población para determinado tiempo. Un parámetro muy utilizado es el número básico de reproducción ( $R_t$ ), este parámetro indica el número promedio de personas que un infectado contagia, si es menor a 1 la enfermedad es controlada y si es mayor a 1 indica que la enfermedad se propaga fácilmente, sin embargo este parámetro es difícil de estimar dado las limitaciones de la información recolectada.

También tuvieron en cuenta modelos como la función exponencial de Gompertz, método de potencia y método de suavizamiento exponencial de Holt-Winters. Estos modelos usan como variable la cantidad de infectados que hay en un día determinado y se le brinda un peso entre 0 y 1 para saber su importancia. Los modelos matemáticos y

estadísticos cumplen con la finalidad de informar y de prever el comportamiento futuro de las variables de interés. Sin embargo, ninguno de estos es infalible.

Debido a la expansión mundial del COVID-19 se han vuelto esenciales los sistemas de información geográfica (SIG) y métodos que incluyen el mapeo en tiempo real de casos de contagio, mapeo predictivo de riesgos y trazado y mapeo de trayectorias y contactos a través del espacio y el tiempo. Estos mecanismos son vitales para la comprensión de la fuente de la enfermedad, dinámica y epidemiología para responder efectivamente, por eso en Kamel y Geraghty ([14]) vemos los dashboards más significativos que se han creado.

El primer dashboard implementado para visualizar la expansión del COVID-19 fue realizado por el Centro de la Universidad de Johns Hopkins para Ciencia e Ingeniería de Sistemas ([9]), presentado en la figura ([1]), fue publicado el 22 de enero de 2020, desarrollado por Lauren Gardner. Este tablero localiza y cuenta infecciones confirmadas, muertes y recuperaciones, sus gráficos detallan el progreso del virus a través del tiempo.

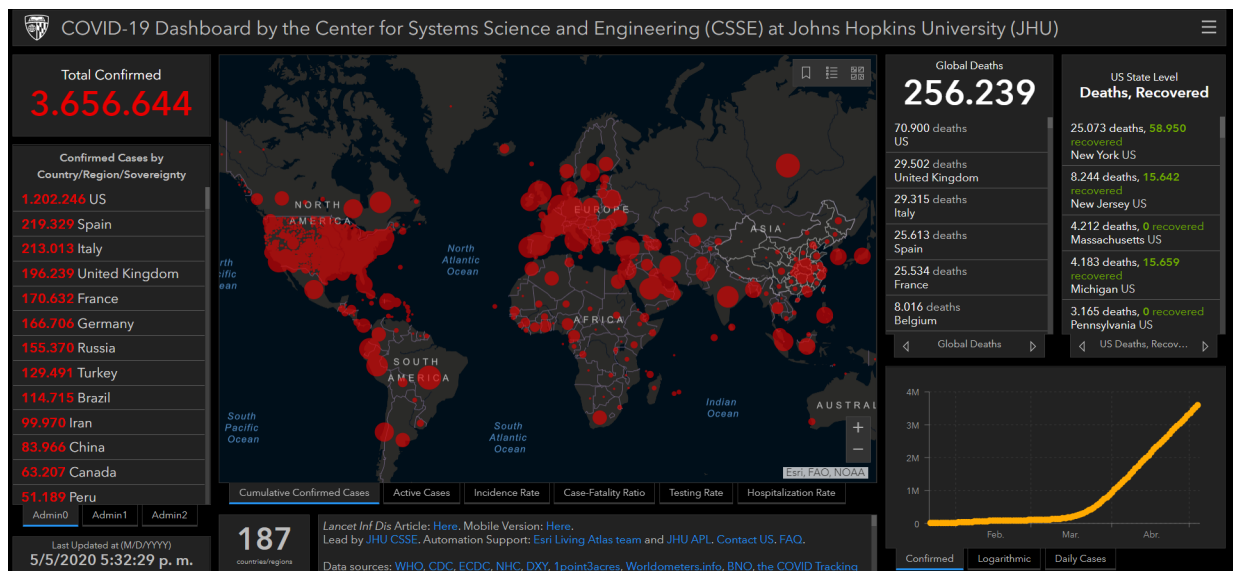


Figura 1: Dashboard de la Universidad Johns Hopkins. Fecha de captura: 5 de Mayo de 2020

El 26 de enero de 2020 la Organización Mundial de Salud ([12]) creó su propio tablero de operaciones, presentado en la figura ([2]), en el cual también mapea y enumera los casos de coronavirus y el número total de muertes por país y provincia china. El panel de la OMS incluye una curva epidémica que muestra los casos por fecha de notificación, esto sirve para conocer la progresión del brote.

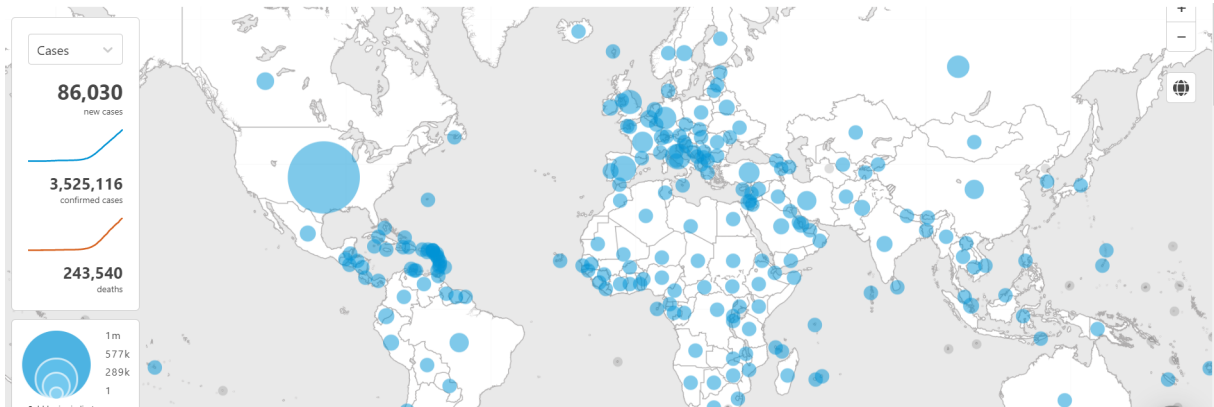


Figura 2: Dashboard de la Organización Mundial de la Salud. Fecha de captura: 5 de Mayo de 2020. Tomado de (16)

HealthMap (13) la cual fue fundada en 2006 y es dirigida por un equipo de investigadores del Bolton Children's Hospital en Estados Unidos, a diferencia de los dos anteriores construye su dashboard, como se muestra en la figura (3), basado en los datos obtenidos por medios de comunicación, redes sociales, alertas validadas por la OMS. Estos datos los obtienen utilizando técnicas de aprendizaje automático y procesamiento de lenguaje natural para examinar informes a través del mundo.

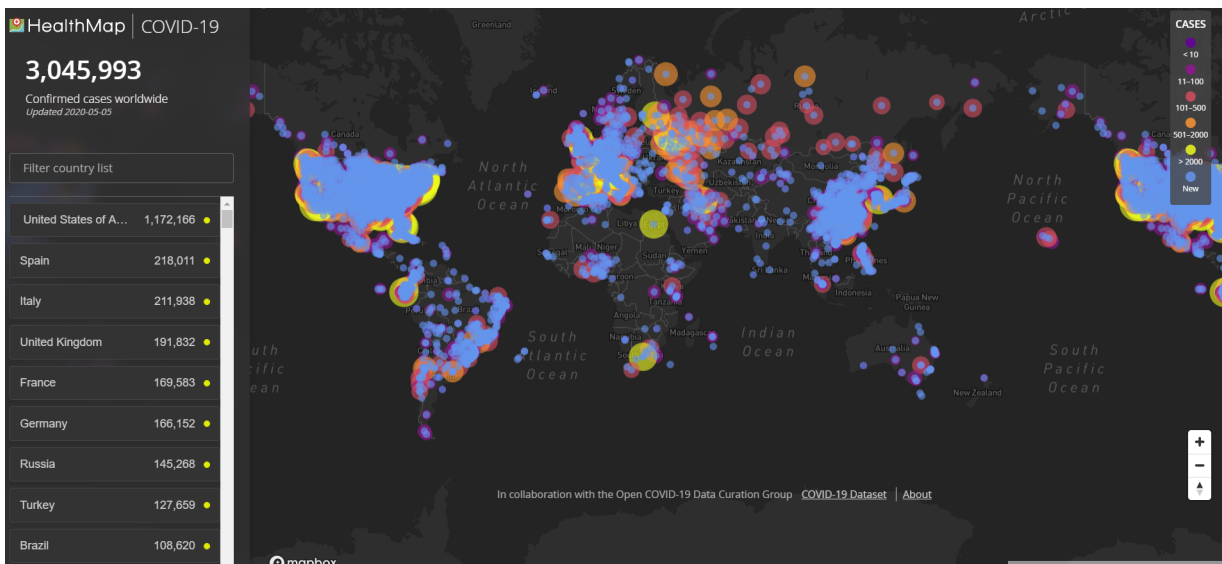


Figura 3: Dashboard de la HealthMap. Fecha de captura: 5 de Mayo de 2020. Tomado de (17)

## 3.2. MARCO TEÓRICO

Actualmente los pronósticos han contado con una atención mayor en los últimos años debido al desarrollo de técnicas más complejas las cuales se han dado gracias a la alta capacidad de computo de los equipos actuales permitiendo la capacidad de análisis de datos para fines de pronóstico lo cual hoy en día es de suma importancia para que las empresas tomen decisiones que les pueda llegar a favorecer a futuro.

Cabe resaltar que el ser humano posee un conocimiento único aprendido a través de los años, dándole la capacidad de crear un juicio a través del instinto el cual es respaldado por técnicas de pronóstico cuantitativas.

### Definición

Para continuar con el estudio es de suma importancia tener claro el concepto de pronóstico, el cual según los autores Everett y Ebert (26) se define como “El proceso de estimación de un acontecimiento futuro proyectando hacia el futuro datos del pasado. Los datos del pasado se combinan sistemáticamente en forma predeterminada para hacer una estimación del futuro” (1991, p. 84).

Además su importancia es trascendental para las empresas puesto que como lo menciona John E. Hanke y Arthur G. Reitsch (11) “las organizaciones operan en una atmósfera de incertidumbre y, a pesar de este hecho, se deben tomar decisiones que afectan el futuro de la organización” por esto, contar con herramientas de pronóstico que ayuden a tener un soporte confiable a la hora de tomar una decisión trascendental para la empresa es importante ya que de esta manera se puede por ejemplo: reducir pérdidas grandes a la empresa permitiéndole estar vigente en el mercado competitivo actual y optimizar la distribución de recursos.

### Selección del método de pronóstico

Para seleccionar el método de pronóstico adecuado de acuerdo al objetivo planteado es necesario tener claro algunos aspectos como: el nivel de detalle, es decir, si el pronóstico necesita detalles específicos o si necesita el estado futuro de algún factor global; si nuestro pronóstico es a corto o largo plazo y el grado apropiado tanto de los métodos cualitativos como de los cuantitativos. Además, es de suma importancia escoger métodos que faciliten la toma de decisiones de las personas que solicitan el pronóstico.



Figura 4: Pasos a seguir en la elaboración de un pronóstico

## Pasos a seguir en la elaboración de un pronóstico

La aceptación de que las técnicas de pronóstico funcionan sobre datos pasados conlleva a la identificación de cuatro pasos en el proceso:

1. **La recopilación de datos:** Según Ferrel y Hartline (5) se requiere tener los datos pertinentes y además que estos sean correctos, es de gran importancia este paso puesto que los demás se realizan con los datos obtenidos. Para recopilar los datos de manera efectiva tenemos las siguientes técnicas:
  - **Observación directa:** El investigador registra las conductas de las personas u objetos en entornos naturales. La observación permite al analista determinar qué se está haciendo, cómo se está haciendo, cuánto tiempo toma, dónde y por qué se hace.
  - **Sesiones de grupo:** El investigador modera un panel de discusión para analizar un tema específico. Es un medio para obtener información acerca de un tema en particular.
  - **Encuestas:** El investigador pide a los encuestados responder una serie de preguntas. Puede ser de manera digital o física.
  - **Experimentos:** El investigador expone a diferentes tratamientos mientras controla las variables extrañas. Permite hacer un análisis y síntesis para entender por qué los productos se desempeñan en la forma que lo hacen.
2. **Reducción o condensación de datos:** Consiste en la reducción de datos dado que solo una cantidad de estos son relevantes en el pronóstico y si no se eliminan pueden reducir la precisión de los resultados.
3. **La construcción del modelo:** En este paso se proponen una serie de modelos que modelan la información y permiten realizar un buen pronóstico. Además, se realiza una estimación de los modelos con una parte de la información (aproximadamente el 80 %) para comparar los pronósticos generados por los diferentes modelos con la información faltante y de esta manera escoger el que mejor se ajuste a la serie a partir de indicadores estadísticos.

4. **Extrapolación del modelo:** En esta etapa se valida el mejor modelo y este es utilizado para realizar finalmente el pronóstico requerido, con sus indicadores de proximidad.

## Medición del error en el pronóstico

Debido a que las técnicas cuantitativas de pronóstico generalmente implican series de tiempo de datos, se desarrolló las siguiente notación matemática donde:

- $Y_t$ : Valor de una serie en el periodo  $t$ .
- $\widehat{Y}_t$ : Valor del pronóstico para  $Y_t$

con  $t = 1, \dots, n$  formando la serie  $Y_1, Y_2, \dots, Y_t, \widehat{Y}_{t+1}, \widehat{Y}_{t+2}, \dots, \widehat{Y}_{t+m}$ .

La mayoría de mediciones para calcular los errores de diferentes técnicas de pronóstico implementan un residual, el cual corresponde a la diferencia entre un valor real y su respectivo valor de pronóstico. Para calcular el residual de cada periodo de pronóstico se utiliza la siguiente ecuación:

$$e_t = Y_t - \widehat{Y}_t. \quad (1)$$

Donde:

- $e_t$ : Error del pronóstico en el periodo  $t$ .
- $Y_t$ : Valor real en el periodo  $t$ .
- $\widehat{Y}_t$ : Valor del pronóstico en el periodo  $t$ .

Entre los métodos para evaluar las diferentes técnicas de pronóstico están:

- **Desviación Absoluta de la Media (DAM):** Mide la precisión de un pronóstico a través de la magnitud de los errores de pronóstico, es decir, los valores absolutos de cada error. Este método es de gran utilidad cuando el analista desea medir el error de pronóstico en las mismas unidades de la serie original.

$$DAM = \frac{\sum_{t=1}^n |Y_t - \widehat{Y}_t|}{n} \quad (2)$$

- **Error medio cuadrático (EMC):** Este método penaliza los errores mayores de pronóstico puesto que eleva cada error al cuadrado, lo cual es útil cuando se requiere una técnica que produzca errores moderados en lugar de una que genere

errores pequeños, pero que ocasionalmente arroje algunos valores extremadamente grandes.

$$EMC = \frac{\left(\sum_{t=1}^n (Y_t - \widehat{Y}_t)\right)^2}{n} \quad (3)$$

- **Porcentaje de error medio absoluto (PEMA):** Este método proporciona una indicación de que tan grande son los errores de pronóstico comparándolos con los valores reales de la serie. El PEMA suele ser útil cuando la magnitud de la variable de pronóstico es importante en la evaluación de la precisión del pronóstico.

$$PEMA = \frac{\sum_{t=1}^n \frac{|Y_t - \widehat{Y}_t|}{Y_t}}{n} \quad (4)$$

- **Porcentaje medio de error (PME):** Este método se suele utilizar cuando es necesario saber si el método de pronóstico se encuentra sesgado.

Si un enfoque de pronóstico no está sesgado, la ecuación de PME producirá un porcentaje cercano a cero, si el resultado es un porcentaje negativo grande, el método de pronóstico está sobrestimando de manera consistente y si el resultado es un porcentaje positivo grande, el método de pronóstico está subestimando en forma consistente.

$$PME = \frac{\sum_{t=1}^n \frac{(Y_t - \widehat{Y}_t)}{Y_t}}{n} \quad (5)$$

- **Raíz del error cuadrático medio (RMSE):** Este método representa la raíz cuadrada de las diferencias entre los valores previstos y los valores observados. La RMSE es sensible a los valores atípicos dado que los errores mayores tienen un efecto muy grande.

$$RMSE = \sqrt{\frac{\left(\sum_{t=1}^n (Y_t - \widehat{Y}_t)\right)^2}{n}}$$

### 3.2.1. Tipos de pronóstico

Los pronósticos se pueden clasificar de acuerdo con Hanke & Reitsh (Pronósticos en los negocios) (II) como:

- **Pronósticos a largo o corto plazo:** Son utilizados para establecer el curso general de una empresa para un largo periodo y los pronósticos a corto plazo son utilizados para diseñar estrategias para la toma de decisiones inmediata.
- **Micro pronósticos o macro pronósticos:** Los pronósticos se pueden clasificar de acuerdo a su posición en el entorno, es decir, según el grado en que intervienen pequeños detalles vs. grandes valores resumidos.
- **Pronósticos de acuerdo a tendencia:** Pronósticos de tipo cualitativo y cuantitativo.

A continuación, nos enfocaremos en esta última clasificación:

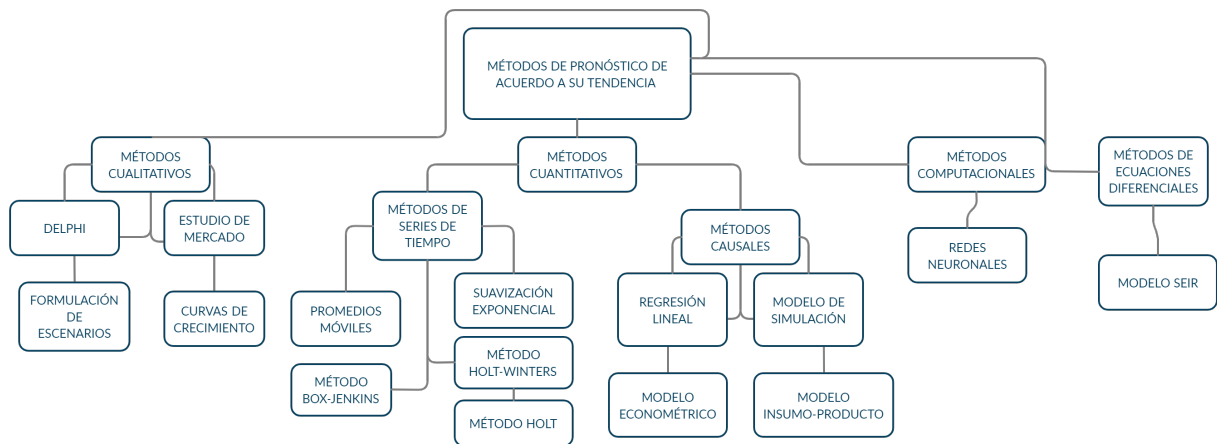


Figura 5: Pronósticos de acuerdo a su tendencia. Construcción propia.

### 3.3. Métodos cuantitativos

#### 3.3.1. Métodos de series de tiempo

Los métodos de análisis de series de tiempo se utilizan para hacer análisis detallados de los patrones históricos de una variable a lo largo del tiempo y proyectarlos hacia el futuro.

Uno de los supuestos básicos de todos los métodos de análisis de series de tiempo es que la variable puede descomponerse en elementos básicos:

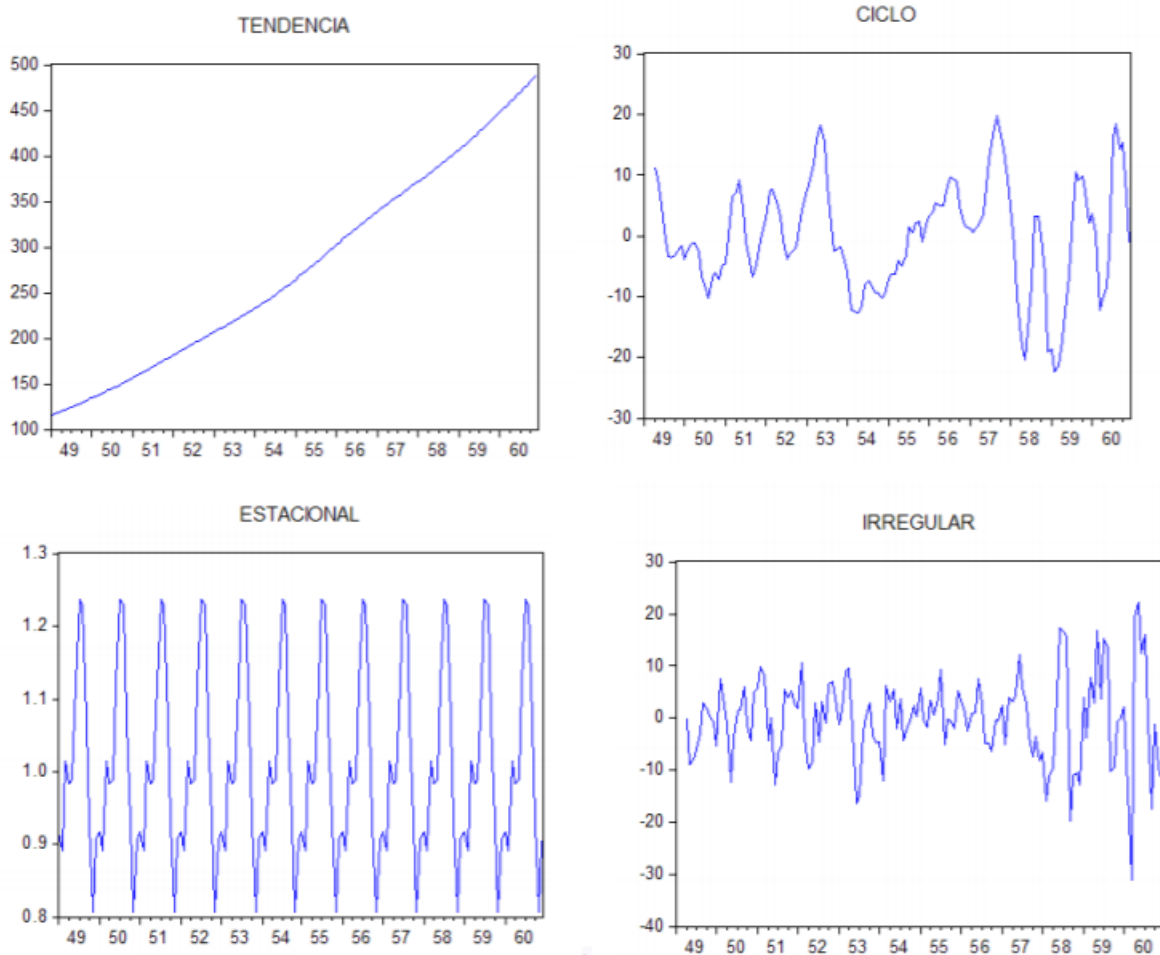


Figura 6: Componentes de una serie de tiempo. Tomado de: Universidad Carlos III (28)

- **Tendencia (T):** Es el componente de largo plazo que representa el crecimiento o disminución en la serie sobre un periodo amplio.
- **Componente cíclico (C):** Es la fluctuación en forma de onda alrededor de la tendencia.
- **Componente estacional (S):** Es un patrón de cambio que se repite a si mismo cada determinado periodo de tiempo el cual puede ser diario, semanal, mensual, anual, entre otros.

- **Componente aleatorio o Irregular (A):** Mide la variabilidad de las series de tiempo después de retirar los otros componentes.

Entre los métodos de series de tiempo se encuentran:

- **Método de promedios móviles:** Los datos históricos se pueden atenuar de maneras diferentes, el objetivo consiste en usar los datos anteriores para así desarrollar un modelo de pronóstico futuro. Un promedio simple se obtiene encontrando la media de todos los valores pertinentes y después utilizar esta media para pronosticar el periodo siguiente. Pero cuando el analista está interesado en las observaciones más recientes, se especifica como conjunto un número de puntos dados y se calcula la media para las observaciones más recientes, recibiendo este enfoque el termino de promedio móvil.

$$M_t = \widehat{Y}_{t+1} = \frac{(Y_t + Y_{t-1} + Y_{t-2} + \dots + Y_{t-n+1})}{n} \quad (6)$$

en donde:

- $M_t$  : Promedio móvil en el periodo  $t$ .
- $\widehat{Y}_{t+1}$  : Valor de pronóstico para el siguiente periodo.
- $Y_t$  : Valor real en el periodo  $t$ .
- $n$  : Número de términos en el promedio móvil.

Se debe tener en cuenta:

- El promedio móvil para el periodo  $t$  es la media aritmética de las  $n$  observaciones más recientes.
- Se asignan ponderaciones iguales a cada observación.
- Al hacerse disponible, cada nuevo punto de datos se incluye en el promedio y se descarta el más antiguo.
- La proporción de respuesta a los cambios en el patrón subyacente de los datos depende del número de periodos  $n$ , que se incluyen en el promedio móvil.
- El modelo de promedio móvil funciona mejor con datos estacionarios<sup>1</sup> puesto que no maneja muy bien la tendencia o estacionalidad.

---

<sup>1</sup>Un proceso es estacionario si su media y su varianza son constantes en el tiempo y si la covarianza entre dos periodos depende del rezago entre ambos periodos. Tomado de:Gobierno de Perú, (29)

### ■ Suavización exponencial

La suavización exponencial es un método utilizado para revisar constantemente una estimación a la luz de experiencias más recientes. El método está basado en el promedio (suavización) de valores anteriores de una serie, haciendo esto de forma decreciente (exponencial). Las observaciones se ponderan, asignando mayor peso a las más recientes. Las ponderaciones se denominan por  $\alpha$  para las más recientes y  $(1 - \alpha)$  para las anteriores.

Se podría pensar en un nuevo pronóstico como el promedio ponderado de la nueva observación y el promedio anterior. Se asigna el valor  $\alpha$  para el nuevo valor y  $1 - \alpha$  para el pronóstico anterior, con  $0 < \alpha < 1$ .

$$\widehat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\widehat{Y}_t \quad (7)$$

donde:

- $\widehat{Y}_{t+1}$  = Valor de pronóstico para el siguiente periodo.
- $\alpha$  = Constante de suavización.
- $Y_t$  = Nueva observación de la serie en el periodo  $t$ .
- $\widehat{Y}_t$  = Valor suavizado de la serie atenuada al periodo  $t - 1$ .

Para interpretar mejor a  $\alpha$  podemos tomar:

$$\widehat{Y}_{t+1} = \widehat{Y}_t + \alpha(Y_t - \widehat{Y}_t) \quad (8)$$

### ■ Métodos de descomposición

Los métodos de descomposición se suelen usar para pronosticar cuando hay estacionalidad en la serie de tiempo o cuando se quiere estudiar la naturaleza de los componentes.

Los datos  $X_t$  pueden expresarse como una función de componente de tendencia  $T_t$ , de una componente estacional  $S_t$  y de un error  $A_t$ , así:

- $T_t$ : Modela el comportamiento de la serie a largo plazo.
- $S_t$ : Modela el comportamiento periódico de la serie.
- $A_t$ : Representa la parte impredecible de la serie.

Cabe resaltar que el componente cíclico no se tiene en cuenta dado que el pronóstico se realiza en el corto plazo.

De esta manera, se encuentran dos métodos de descomposición:

- **Método de descomposición aditivo:** Es apropiado cuando las fluctuaciones estacionales no varían con el nivel de la serie.

$$X_t = T_t + S_t + A_t \quad (9)$$

- **Método de descomposición multiplicativo:** Es apropiado cuando las fluctuaciones estacionales crezcan o decrezcan proporcionalmente con los incrementos o decrementos del nivel de la serie.

$$X_t = (T_t)(S_t)(A_t) \quad (10)$$

#### ■ Método de Holt

Según Izar (27) este modelo añade al modelo de suavizamiento exponencial un componente debido a la tendencia, por esta razón es conocido como suavizamiento exponencial doble. Se presentan dos parámetros importantes como lo son  $\alpha$  la constante de suavizamiento y  $\beta$  la constante de tendencia. La variable pronosticada para el tiempo  $t$  está dada por:

$$\widehat{Y}_t = L_t + pT_t$$

Donde:

- $\widehat{Y}_t$  : Valor pronosticado para el periodo  $t$ .
- $L_t$  : Valor estimado para el periodo  $t$ .
- $T_t$  : Valor de la tendencia para el periodo  $t$ .
- $p$  : Periodos futuros a pronosticar.

El valor estimado para la variable en el periodo  $t$  está dado por la ecuación:

$$L_t = \alpha Y_{t-1} + (1 - \alpha)(L_{t-1} + T_{t-1})$$

Y la tendencia está dada por:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

## ■ Método de Holt-Winters

El método de Holt Winters es un método iterativo que en cada tiempo realiza un pronóstico sobre el comportamiento de una serie. Además, es una versión extendida del método de Holt, puesto que el método de Holt-Winters utiliza así como el método de Holt, un parámetro para el nivel y uno para la tendencia, y en forma adicional se tiene un parámetro más para la estacionalidad, los cuales son generalmente llamados alfa ( $\alpha$ ), beta ( $\beta$ ) y gamma ( $\gamma$ ).

Para la implementación de este método se hace uso de datos históricos para obtener una nueva serie más suave a partir de la cual se realiza una previsión. El método de Holt-Winters se puede presentar de forma multiplicativa o aditiva, para determinar que forma elegir se debe tener en cuenta la tendencia y estacionalidad de los datos, pues si ambas aumentan a lo largo del tiempo entonces tenemos un modelo de Holt-Winters multiplicativo mientras que si no aumentan se presenta un modelo Holt-Winters aditivo.

El método de Holt-Winters utiliza la siguiente estructura:

### Estimación del nivel:

$$N_t = \alpha + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (11)$$

Donde:

- $N_t$ : Valor suavizado del nivel del periodo  $t$ .
- $\alpha$ : Constante de suavizamiento del nivel.
- $N_{t-1}$ : Valor suavizado del nivel del periodo  $t - 1$ .
- $T_{t-1}$ : Estimación de la tendencia en el periodo  $t - 1$ .

### Estimación de la tendencia:

$$T_t = \beta(N_t - N_{t-1}) + (1 - \beta)T_{t-1} \quad (12)$$

Donde:

- $T_t$ : Estimación de la tendencia en el periodo  $t$ .
- $\beta$ : Constante de suavizamiento de la tendencia.
- $N_t$ : Valores suavizados del nivel del periodo  $t$ .
- $N_{t-1}$ : Valores suavizados del nivel del periodo  $t - 1$ .
- $T_{t-1}$ : Estimación de la tendencia en el periodo  $t - 1$ .

### Estimación de la estacionalidad:

$$E_t = \gamma + (1 - \gamma)E_{t-s} \quad (13)$$

Donde:

- $E_t$ : Estimación de la estacionalidad en el periodo  $t$ .
- $\gamma$ : Constante de suavizamiento de la estacionalidad.
- $s$ : Subíndice que representa la longitud de la estacionalidad.
- $E_{t-s}$ : Estimación de la estacionalidad en el periodo  $t - s$ .

### Previsión:

$$F_{t+1} = (N_t + T_t)E_{t-s+1} \quad (14)$$

Donde:

- $F_{t+1}$ : Previsión del periodo  $t$ .
- $N_t$ : Valor suavizado del nivel del periodo  $t$ .
- $T_t$ : Estimación de la tendencia en el periodo  $t$ .
- $s$ : Subíndice que representa la longitud de la estacionalidad.
- $E_{t-s+1}$ : Estimación de la estacionalidad en el periodo  $t - s + 1$ .

Tomado de: Anónimo (8).

### ■ Método de Box-Jenkins

La metodología de los modelos ARIMA según De Lara (6) fue formalizada por Box y Jenkins por lo que también se le conoce como método de Box-Jenkins.

Este enfoque asume que la serie temporal a predecir es generada por un proceso estocástico cuya naturaleza se puede generalizar mediante un modelo. Para estimar una serie ARIMA se requiere de un gran número de observaciones.

Los modelos ARIMA están compuestos por tres partes:

- **Autorregresivo AR(p):** El proceso autorregresivo de orden  $p$  expresa la variable de respuesta  $Y_t$  en función de los valores pasados hasta el retardo  $t - p$ , está dado por la siguiente ecuación:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t$$

donde:

- $a_t \sim N(0, \sigma^2)$ .
- $\phi_t$ : Parámetros pertenecientes a la parte autorregresiva.
- **Integrado (I):** La parte integrada de un modelo ARIMA se refiere a la cantidad de diferenciaciones<sup>2</sup> que se realiza en el modelo para eliminar la estacionalidad.
- **Medias Móviles MA(q):** El modelo de medias móviles de orden finito  $q$  se obtiene de truncar el modelo de medias móviles de orden infinito de la siguiente manera:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

donde:

- $a_t \sim N(0, \sigma^2)$ .
- $\theta_t$  : Parámetros pertenecientes a la parte de medias móviles.

En los modelos ARIMA univariantes se explica el comportamiento de una serie temporal a partir de observaciones pasadas de la propia serie y a partir de los errores pasados de previsión.

La notación compacta de los modelos ARIMA está dada por:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} - (\Delta^d Y_t - Y_t)$$

donde:

- $a_t \sim N(0, \sigma^2)$ .
- $\theta_t$  : Parámetros pertenecientes a la parte de medias móviles.
- $\phi_t$ : Parámetros pertenecientes a la parte autorregresiva.

---

<sup>2</sup>Diferenciación: Método para eliminar la tendencia que consiste en restar a cada valor de la serie el valor anterior. Tomado de: Marín (30)

- $d$  : Cantidad de diferenciaciones requeridas para eliminar la tendencia.
- $\Delta Y_t$  : Variación entre cada par de tiempos consecutivos.

La parte autorregresiva (AR) del modelo es  $\phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p}$ , la parte de medias móviles (MA) es  $-\theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$  y la parte de diferenciación (I) está dada por  $-(\Delta^d Y_t - Y_t)$ .

Los coeficientes de los parámetros  $\phi_0, \phi_1, \phi_p, \theta_1, \theta_q$  son determinados a partir de los datos mediante algún método estadístico consistente.

## Modelos SEIR

Los procesos de transmisión de agentes infecciosos en poblaciones es un objeto de estudio de los modelos epidemiológicos compartimentales, los cuales se basan en sistemas de ecuaciones diferenciales para el movimiento de la población en estados discretos. El modelo SEIR presenta las poblaciones susceptibles a la enfermedad (S), infectados que no transmiten la enfermedad (E), infectados que pueden infectar (I) y recuperados (R). El sistema de ecuaciones diferenciales está dado por:

$$\begin{aligned}
 N &= S + E + I + R \\
 \frac{dS}{dt} &= \mu(N - S) - \beta \frac{SI}{N} - \eta S \\
 \frac{dE}{dt} &= \beta \frac{SI}{N} - (\mu + \sigma)E \\
 \frac{dI}{dt} &= \sigma E - (\gamma + \mu)I \\
 \frac{dR}{dt} &= \gamma I - \mu R + \eta S
 \end{aligned}$$

donde:

- $\mu$  : Tasa de mortalidad en la población no relacionada con la enfermedad.
- $\beta$  : Tasa efectiva de contacto.
- $\eta$  : Tasa de vacunación (si existe la vacuna).
- $\sigma$  : Tasa a la cual una persona expuesta se convierte en infecciosa.
- $\gamma$  : Tasa a la cual un infectado se recupera.

El modelo asume que la población es constante, que la tasa de letalidad es baja y todos los infectados adquieren inmunidad. El número reproductivo ( $R_0$ ) representa el número de individuos que contagia cada infectado y está dado por la siguiente ecuación:

$$R_0 = \frac{\sigma\beta}{(\sigma + \mu)(\gamma + \mu)}$$

Si el valor es mayor a 1 significa que cada persona infectada contagia a más de una persona y la enfermedad es capaz de propagarse e infectar a la población susceptible.

### 3.4. RESULTADOS OBTENIDOS

#### 3.4.1. Presentación de variables

Para el estudio se tomaron los datos recogidos por el Instituto Nacional de Salud (25) en el portal de Datos Abiertos Colombia, las variables que se presentan son las siguientes:

VARIABLES	DESCRIPCIÓN
<i>CIUDAD DE UBICACIÓN</i>	Ciudad de residencia del paciente registrado
<i>CONFIRMADOS</i>	Cantidad de confirmados totales
<i>FECHA DE NOTIFICACIÓN</i>	Fecha de notificación del caso a INS
<i>FIS</i>	Fecha de inicio de síntomas
<i>FECHA DIAGNOSTICO</i>	Fecha de confirmación por laboratorio
<i>FECHA DE MUERTE</i>	Fecha de muerte del paciente
<i>FECHA REPORTE WEB</i>	Fecha de publicación en sitio web
<i>CÓDIGO DIVIPOLA</i>	Código numérico asignado a cada departamento
<i>DEPARTAMENTO</i>	Departamento de ubicación del paciente
<i>UBICACIÓN</i>	Tipo de atención brindada al paciente
<i>EDAD</i>	Edad del paciente
<i>SEXO</i>	Sexo del paciente
<i>TIPO</i>	Tipo de contagio del paciente
<i>ESTADO</i>	Estado de salud actual del paciente
<i>PAÍS DE PROCEDENCIA</i>	País donde se contagió el paciente
<i>FECHA RECUPERADO</i>	Fecha de recuperación

Cuadro 1: Variables. Tomado de: INS (25)

### 3.4.2. Análisis univariado

#### Atención <sup>3</sup>

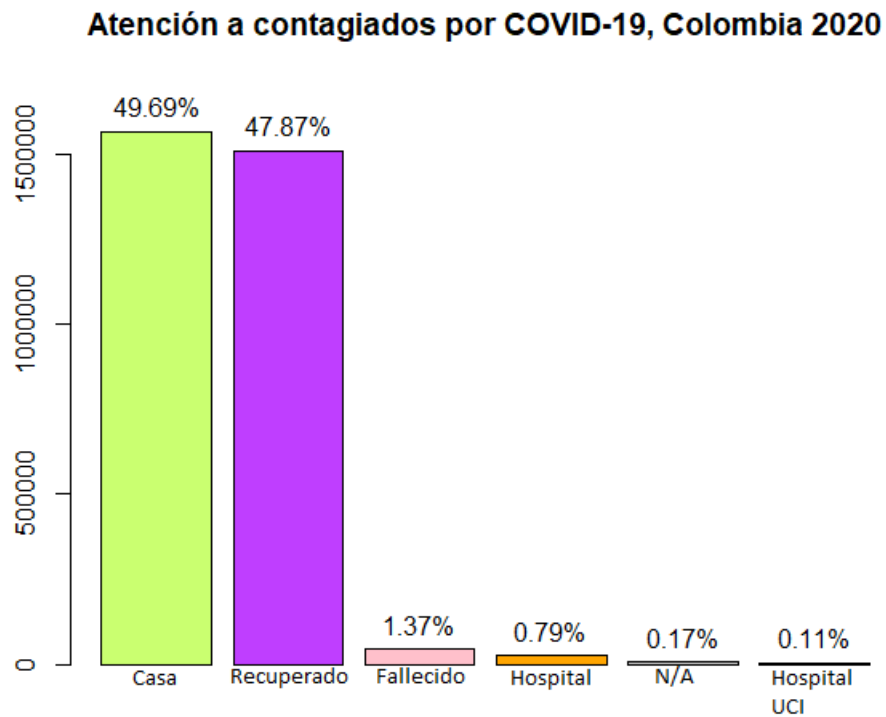


Figura 7: Tipo de atención. Tomado de: INS (25)

El tipo de atención brindada a los pacientes se distribuye en: Casa (49.69%), Fallecido (1.37%), Hospital (0.79%), Hospital UCI (0.11%), Recuperado (47.87%), se registran un 0.17% de los casos en los cuales no se tiene información clara en donde están recibiendo atención.

<sup>3</sup>La fecha de corte para el análisis univariado y bivariado descriptivo es el 31 de diciembre de 2020.

## Tipo

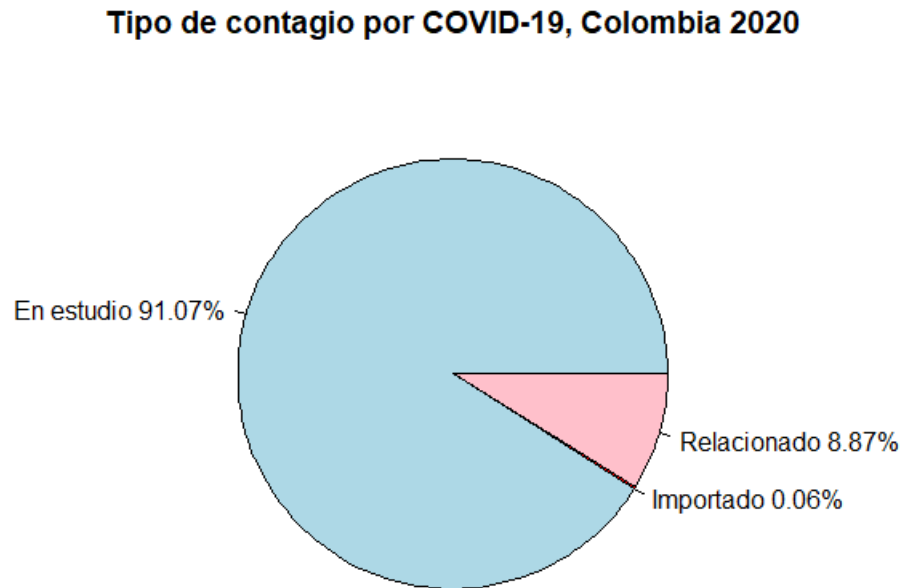


Figura 8: Tipo de contagio. Tomado de: INS (25)

Los casos de contagios de COVID-19 registrados en Colombia son de dos tipos: importados y relacionados, sin embargo la mayoría de los casos que se presentan siguen en estudio para saber con certeza cuál es su procedencia. Cabe resaltar que dado a la rápida decisión de cerrar las fronteras aéreas, terrestres y marítimas la cantidad de casos importados es muy baja.

## Edad

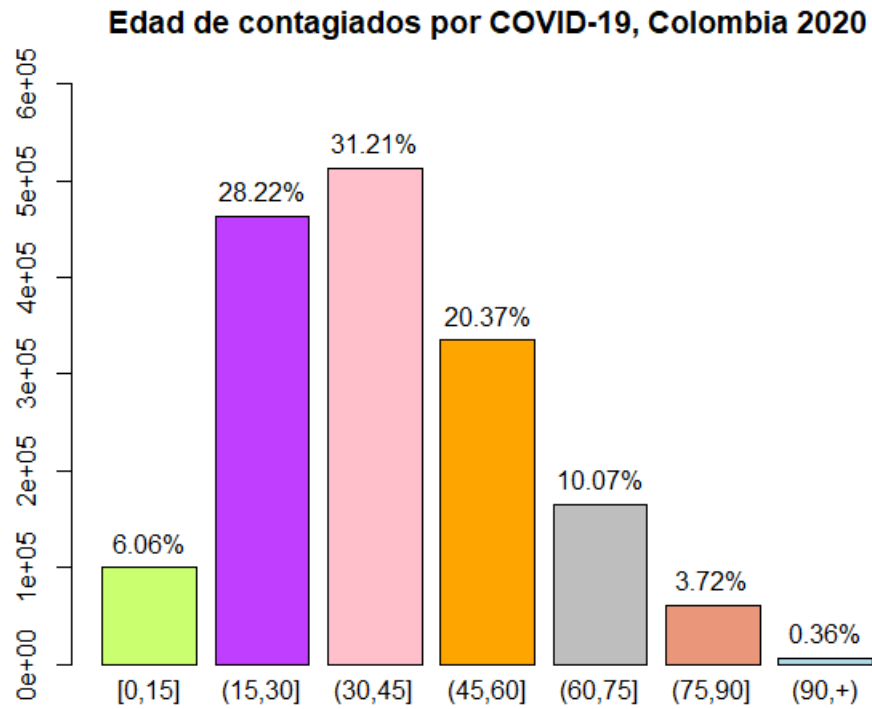


Figura 9: Contagiados por rango de edad. Tomado de: INS (25)

En la figura (9) podemos ver que entre los 15 y los 45 años años se presenta el 59.43% de los casos de contagio en Colombia. Igualmente a medida que la edad aumenta los casos de contagio disminuyen.

## Departamentos

En los cuadros (2) y (3) se registran los contagiados y fallecidos por departamento, vemos que en los 32 departamentos junto al distrito capital de Bogotá se han presentado casos de contagio, concentrándose en los departamentos principales de Colombia como son: Bogotá (896852), Antioquia (505581), Valle del Cauca (246513), Atlántico (181291).

Departamento	Contagiados	Departamento	Contagiados
VAUPES	1124	BOYACA	27818
VICHADA	1134	NARIÑO	30116
GUAINIA	1210	CALDAS	31681
GUAVIARE	1921	CESAR	32391
SAN ANDRES	2421	META	33057
AMAZONAS	3109	RISARALDA	33278
ARAUCA	4417	HUILA	33593
CHOCO	4674	NORTE DE SANTANDER	38054
PUTUMAYO	5600	TOLIMA	42623
CASANARE	8564	BOLIVAR	50491
LA GUAJIRA	13559	SANTANDER	64360
CAQUETA	14365	CUNDINAMARCA	65085
SUCRE	16690	ATLANTICO	89946
CAUCA	18086	VALLE DEL CAUCA	132775
QUINDIO	22896	ANTIOQUIA	255958
MAGDALENA	23528	SANTAFE DE BOGOTA D.C	459136
CORDOBA	27647		

Cuadro 2: Cantidad de contagiados por departamento. Tomado de: INS (25)

Departamento	Fallecidos	Departamento	Fallecidos
VAUPES	12	META	741
VICHADA	14	RISARALDA	752
GUAINIA	20	NARIÑO	974
GUAVIARE	37	CESAR	995
SAN ANDRES	39	MAGDALENA	1072
AMAZONAS	125	BOLIVAR	1092
ARAUCA	134	HUILA	1125
CHOCO	169	TOLIMA	1265
CASANARE	182	CORDOBA	1678
PUTUMAYO	244	CUNDINAMARCA	1712
CAUCA	502	NORTE DE SANTANDER	2010
CAQUETA	530	SANTANDER	2383
LA GUAJIRA	541	ATLANTICO	3433
BOYACA	614	VALLE DEL CAUCA	4274
CALDAS	640	ANTIOQUIA	4786
QUINDIO	652	SANTAFE DE BOGOTA D.C	9788
SUCRE	678		

Cuadro 3: Cantidad de fallecidos por departamento. Tomado de: INS (25)

Podemos ver que los departamentos con mayor cantidad contagio presentan un gran número de muertes.

### 3.4.3. Análisis descriptivo bivariado

#### Cantidad de muertos por edad

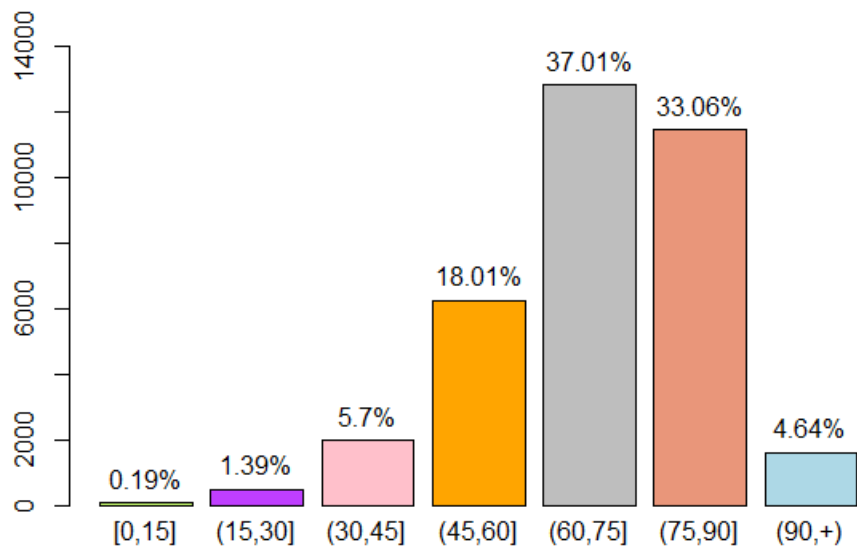


Figura 10: Cantidad de muertos por edad. Tomado de: INS (25)

De la figura (10) podemos concluir que la mayoría de las muertes se han presentado en adultos de la tercera edad concentrándose la mayoría de estas muertes entre los 60 y 90 años.

## Cantidad de recuperados por edad

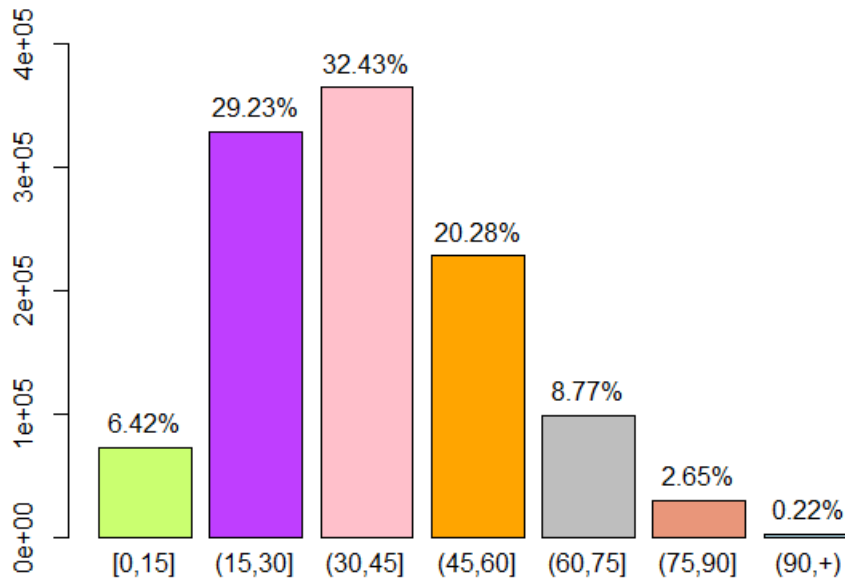


Figura 11: Cantidad de recuperados por edad. Tomado de: INS (25)

Por otro lado, la mayoría de recuperados se encuentran entre los 15 y los 45 años.

## Contagios por edad y sexo

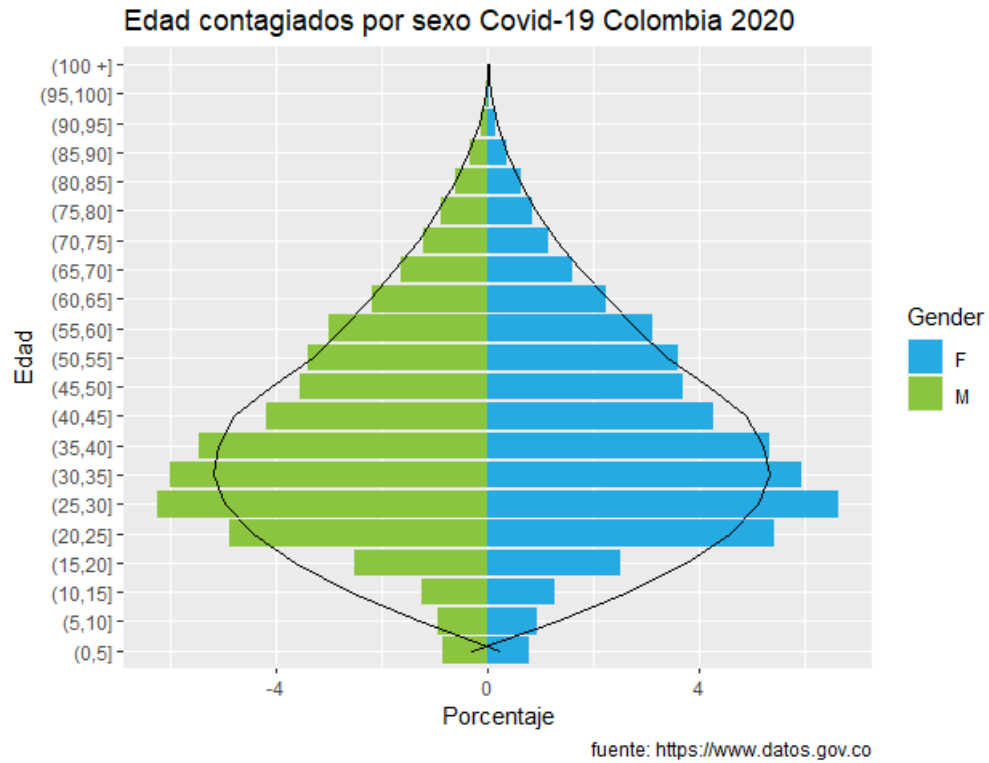


Figura 12: Pirámide de contagios por edad y sexo. Tomado de: INS (25)

Podemos ver que la distribución por sexo es similar en cuanto a los contagios, además tanto en hombres como en mujeres vemos que la mayoría de los contagiados se encuentran entre los 20 y los 40 años.

## ANÁLISIS DE SERIES DE TIEMPO

Es importante determinar cuales fueron las fechas importantes en Colombia durante el tiempo de la pandemia para de esta manera poder interpretar de manera adecuada la curva de contagios durante este año.

En primer lugar, presentamos la curva de contagios desde el 6 de marzo de 2020 hasta el 31 de diciembre de 2020.

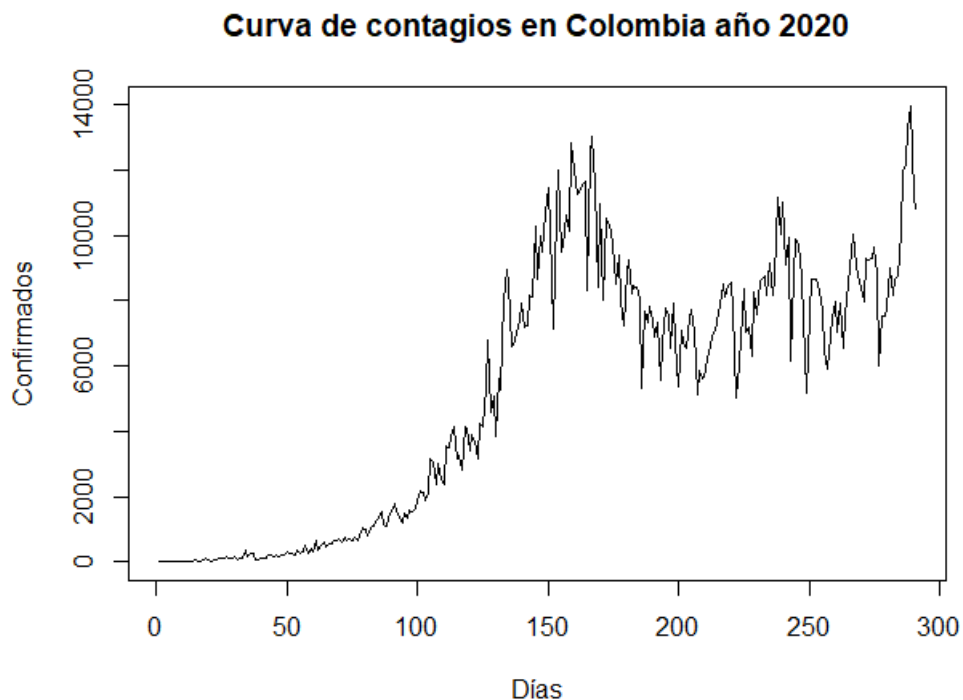


Figura 13: Curva de contagiados en Colombia durante el año 2020

El día 6 de marzo de 2020 se presentó el primer contagio por COVID-19 en la ciudad de Bogotá.

1. **24 de marzo:** Se decreta cuarentena en todo el territorio colombiano.
2. **6 de abril:** Se extendió por primera vez la cuarentena.
3. **5 de mayo:** Se extiende por segunda vez la cuarentena pero con reapertura gradual de algunos sectores productivos.
4. **19 de junio:** Se realizó el día sin IVA con la intención de reactivar la economía, sin embargo se presentaron una gran cantidad de aglomeraciones en toda Colombia.

5. **28 de julio:** Se extendió por tercera vez la cuarentena, pero en esta ocasión las poblaciones con menor cantidad de contagiados podían realizar la reapertura gradual de todas sus actividades económicas.
6. **18 de agosto:** Se declara la semana del pico de la primera ola de la pandemia.
7. **30 de agosto:** Fin de la cuarentena obligatoria en Colombia.
8. **9 y 10 de septiembre:** En estas fechas se presentan una ola de protestas debido al fallecimiento de Javier Ordóñez a manos de la policía.
9. **1 de diciembre:** Inicio de la temporada navideña.
10. **27 de diciembre:** Partido América vs Santa Fe. Final del fútbol colombiano.

### Curva de contagios en Colombia año 2020

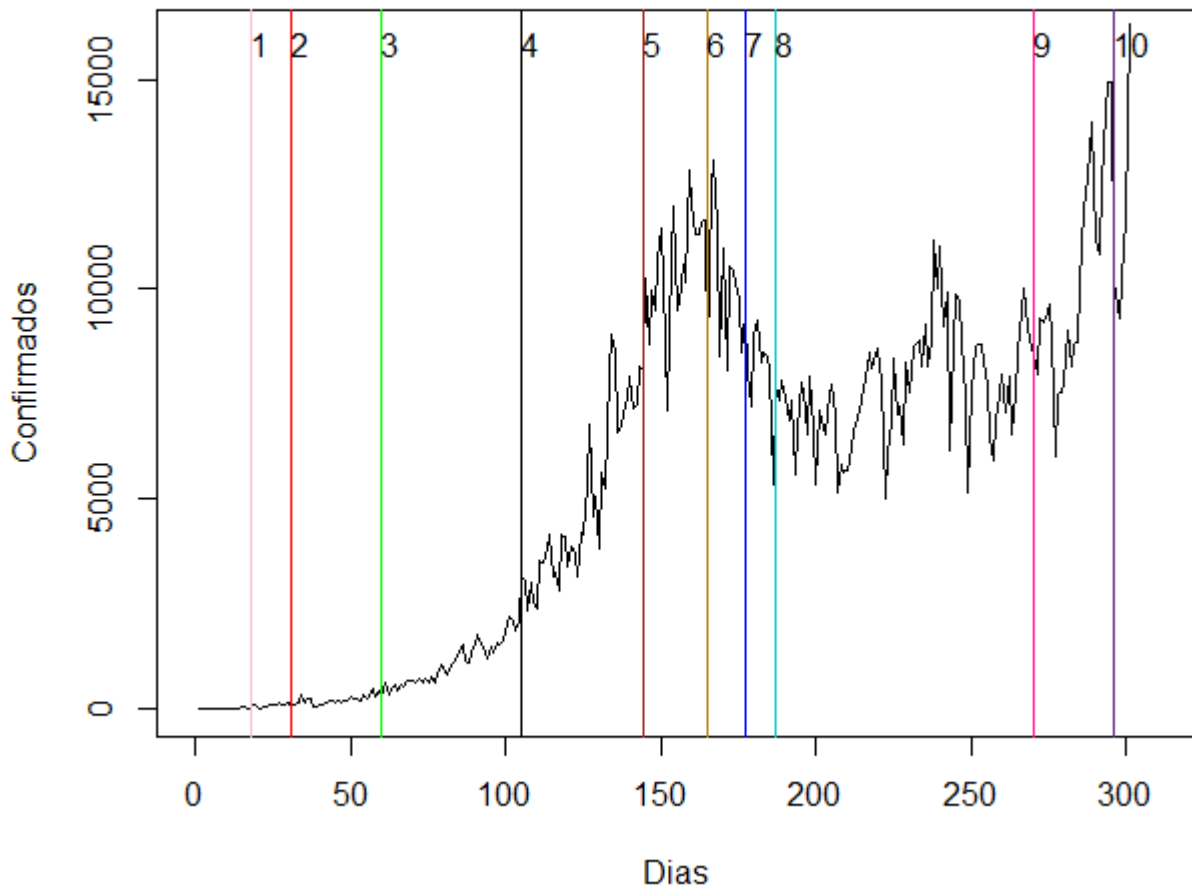


Figura 14: Fechas destacadas durante la pandemia

Como podemos ver en la gráfica, a partir del 19 de junio que se presentó el día sin IVA incrementaron de manera notable los casos confirmados de COVID-19, esto nos indica que las aglomeraciones que sucedieron ese día fueron un foco importante de contagios. Vemos que después de cada extensión de cuarentena se presenta un incremento del número de casos diarios de contagiados. Sin embargo, podemos decir que fue una buena decisión por parte del gobierno mantenerla de manera obligatoria dado que de no ser así se podía haber presentado un número mayor de contagios por COVID-19. Por último, luego del fin de la cuarentena la cantidad de contagios por COVID-19 se mantuvo estable, no obstante a partir del 1 de diciembre vemos un incremento mayor en los contagios diarios debido a las festividades decembrinas.

### 3.5. Generación de Modelos

Después de realizar un análisis de las variables suministradas por la base de datos del ministerio de salud procedemos a generar diferentes modelos aplicados a series de tiempo para posteriormente compararlos y saber cual es el más eficiente para pronosticar al día 31 de diciembre de 2020 la cantidad de contagios en Colombia. Finalmente, la serie de tiempo seleccionada para trabajar los modelos fue la serie dada por los casos de contagios confirmados diarios que será la variable utilizada en el estudio. Para poder generar los modelos primero se realizó un estudio de la estacionalidad y tendencia de la serie. Los modelos utilizados son Holt-Winters, Holt, Medias Móviles y ARIMA. Para la generación de estos modelos se utilizó el lenguaje estadístico R y Rstudio <https://rstudio.com/> Versión: 1.2.5033.

Para todas las gráficas de los pronósticos se presenta en color negro los datos originales obtenidos del Instituto Nacional de Salud, en rojo la estimación de los datos anteriores dada por cada modelo y en azul podemos ver el pronóstico hasta el 31 de diciembre de 2020.

El pronóstico se realizó basado en los datos tomados del Instituto Nacional de Salud desde el día 6 de marzo de 2020 hasta el 21 de diciembre de 2020 y se busca pronosticar la cantidad de contagios diarios que se presentarán del 22 al 31 de diciembre de 2020.

#### Modelo Holt-Winters

Para generar el modelo Holt-Winters en R se utilizó la función *HoltWinters* del paquete *stats*, la cual recibe como parámetro la serie de tiempo y se debe especificar si se utiliza el método multiplicativo o aditivo, en este caso dado que en el estudio previo de la serie de tiempo vemos que tiene una tendencia creciente entonces se utiliza el método multiplicativo. Luego de esto se utiliza la función *predict* para pronosticar hasta el día 31 de diciembre de 2020 la cantidad de contagiados que tendrá Colombia, los resultados son los siguientes:

Fecha	Pronóstico
22 de diciembre	11527.72
23 de diciembre	11547.03
24 de diciembre	11566.33
25 de diciembre	11585.64
26 de diciembre	11604.94
27 de diciembre	11624.25
28 de diciembre	11643.55
29 de diciembre	11662.86
30 de diciembre	11682.16
31 de diciembre	11701.47

Cuadro 4: Pronóstico modelo Holt-Winters

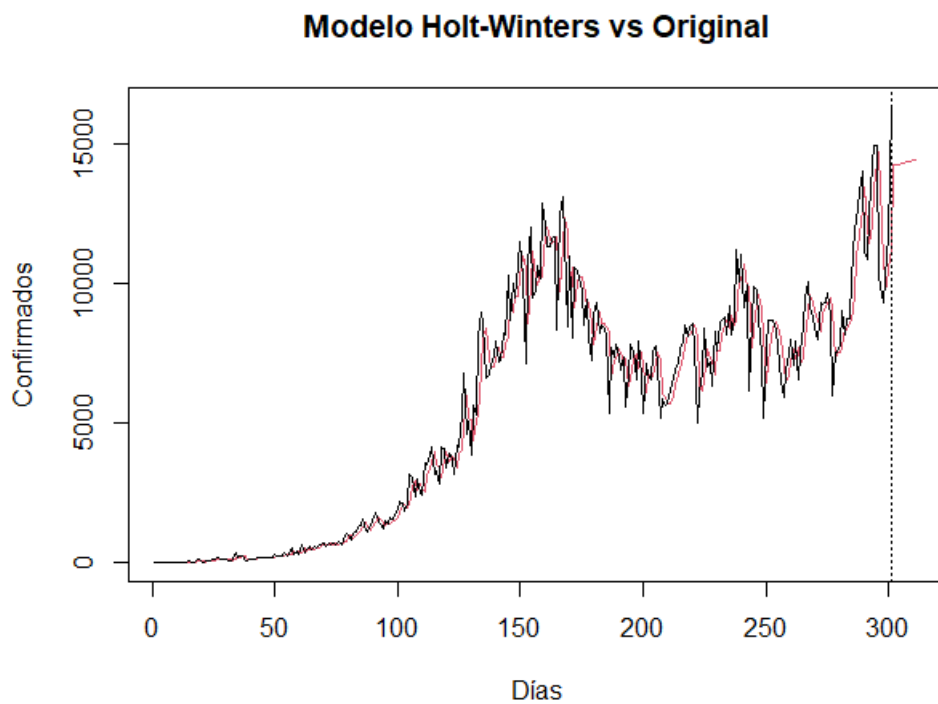


Figura 15: Pronóstico modelo Holt-Winters

En color rojo podemos apreciar los valores que estima el modelo, como se puede apreciar el modelo es muy cercano a los valores reales de contagios.

## Modelo Holt

Para el modelo Holt se utilizó la función *Holt* del paquete *aTSA* de R, las predicciones con este modelos son las siguientes:

Fecha	Pronóstico
22 de diciembre	12447.21
23 de diciembre	12839.93
24 de diciembre	13232.64
25 de diciembre	13625.36
26 de diciembre	14018.08
27 de diciembre	14410.80
28 de diciembre	14803.52
29 de diciembre	15196.24
30 de diciembre	15588.96
31 de diciembre	15981.67

Cuadro 5: Pronóstico modelo Holt

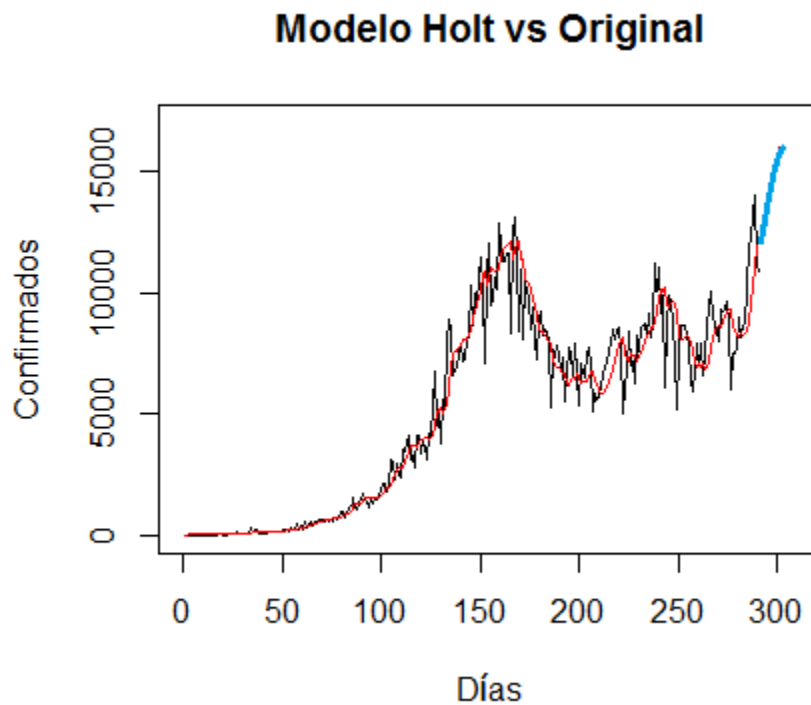


Figura 16: Pronóstico modelo Holt

El modelo Holt a diferencia del modelo de Holt-Winters vemos que presenta diferencias

notables respecto a las estimaciones que realiza para los datos anteriores.

### Modelo de Medias Móviles

Para el caso de medias móviles se analizó el caso con 3, 7, 9 y 14 medias, para elegir el que mejor aproximara se tuvieron en cuenta los errores de esviación Absoluta de la Media (DAM), Error Medio Cuadrático (EMC), Raíz del Error Cuadrático Medio (RMSE).

MEDIAMOV	DAMMA	RMSEMA	EMCMA
3	349.7	556.31	37.65
7	475.93	757.74	77.61
9	512.16	799.08	9058.19
14	517.34	801.84	119545.25

Cuadro 6: Errores de medias móviles

Como podemos ver, para 3 medias móviles se presenta un error menor. Por esta razón, trabajaremos con este modelo para realizar el pronóstico.

Fecha	Pronóstico
22 de diciembre	12119.95
23 de diciembre	12227.86
24 de diciembre	12335.76
25 de diciembre	12443.66
26 de diciembre	12551.57
27 de diciembre	12659.47
28 de diciembre	12767.38
29 de diciembre	12875.28
30 de diciembre	12983.18
31 de diciembre	13091.09

Cuadro 7: Pronóstico de medias móviles

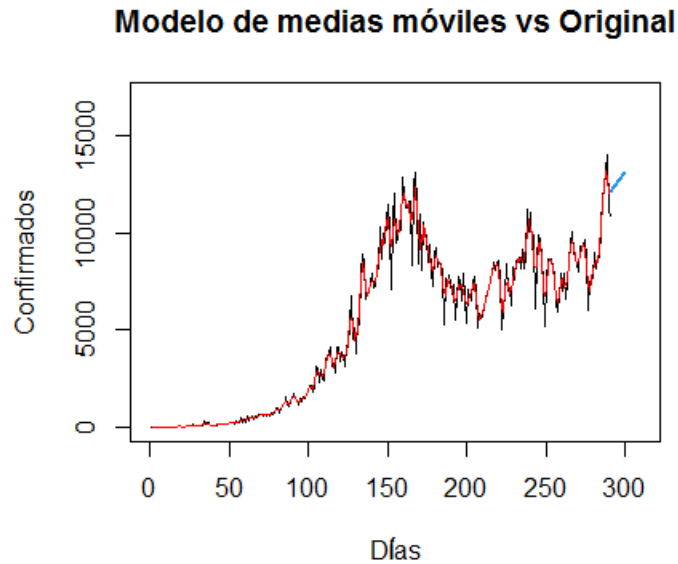


Figura 17: Pronóstico modelo Medias Móviles para  $n = 3$

En el caso de las medias móviles vemos que ambas gráficas son muy parecidas, sin embargo se presenta un desfase entre una gráfica y otra dado que el modelo de medias móviles elimina datos iniciales y datos finales para realizar el pronóstico.

### Modelo ARIMA

Para el caso del modelo ARIMA analizamos primero si tiene tendencia, en nuestro caso mediante el comando *ndiffs* del paquete *forecast* de R podemos concluir que la tiene, para eliminarla realizamos una diferencia la cual consta de realizar la resta entre cada par de datos consecutivos para crear una nueva serie.

Para conocer el grado del modelo Autorregresivo (AR) y el modelo de Medias móviles (MA) debemos realizar el gráfica de autocorrelación y autocorrelación parcial, las gráficas son las siguientes:

### Función de autocorrelación

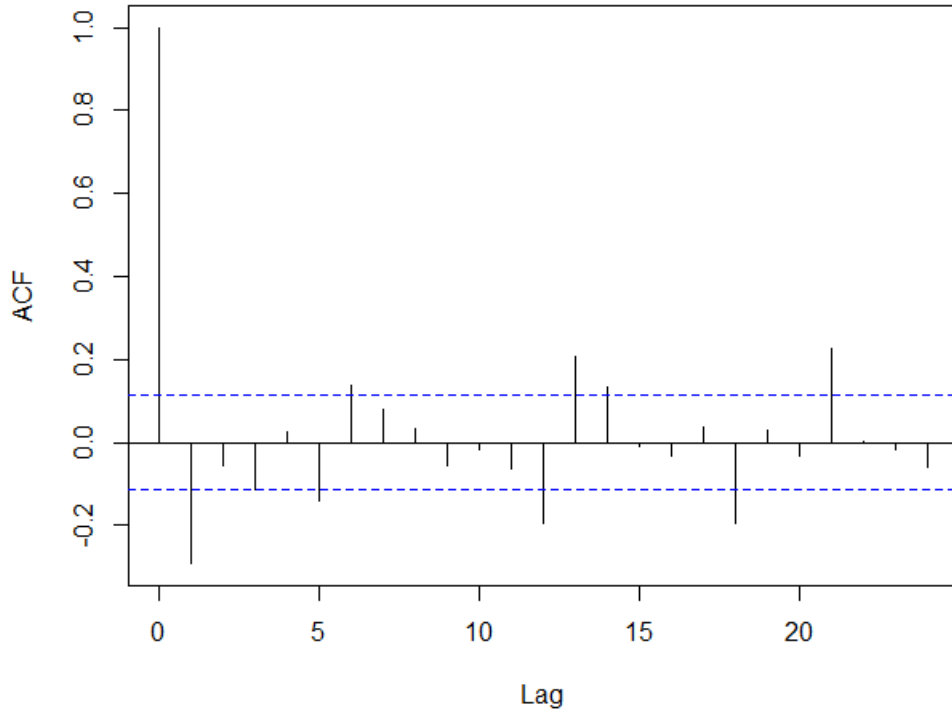


Figura 18: Función de autocorrelación

En este caso podemos ver que el grado de autorregresión es 1 dado que en la parte superior solo aparece una recta mayor a 0.1. Para el modelo de medias móviles tenemos:

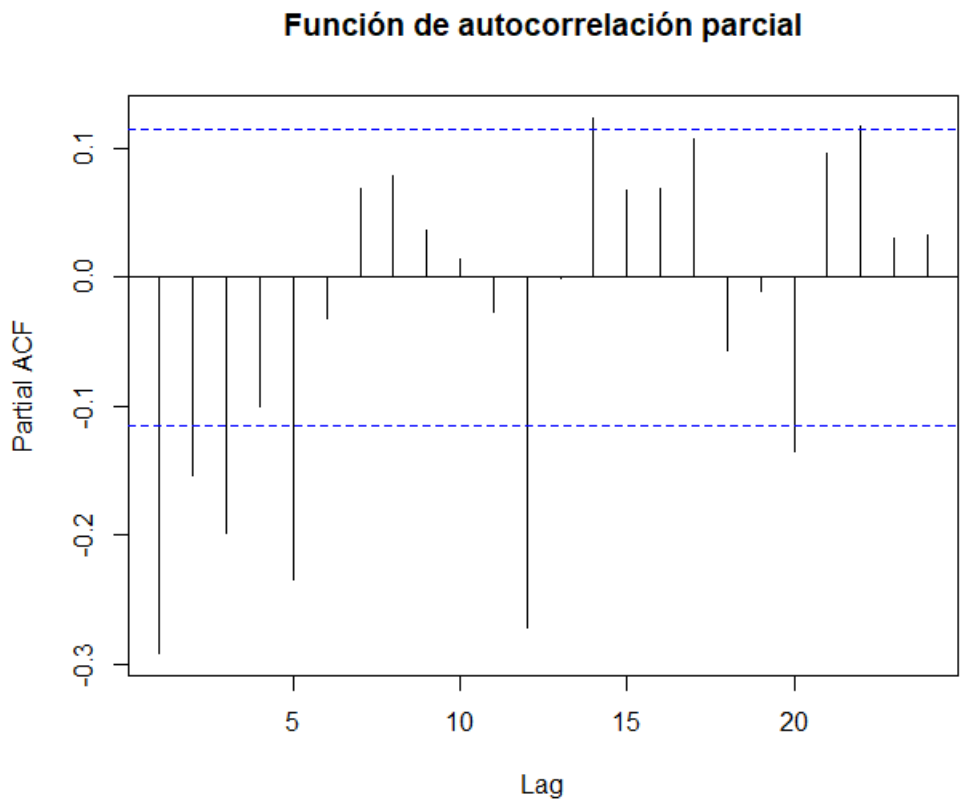


Figura 19: Función de autocorrelación parcial

La función de autocorrelación parcial nos muestra que el grado del modelo de medias móviles es 3. Con esto el modelo arima sería de la forma  $ARIMA(1, 1, 3)$ .

Mediante el comando *forecast* pronosticamos basado en nuestro datos anteriores usando el modelo  $ARIMA(1, 1, 3)$ , el gráfico es el siguiente:

Fecha	Pronóstico
22 de diciembre	11532.39
23 de diciembre	11843.00
24 de diciembre	11692.96
25 de diciembre	11854.31
26 de diciembre	11822.29
27 de diciembre	11904.44
28 de diciembre	11921.41
28 de diciembre	12767.38
29 de diciembre	11974.71
30 de diciembre	12008.11
31 de diciembre	12052.26

Cuadro 8: Pronóstico mediante modelo ARIMA

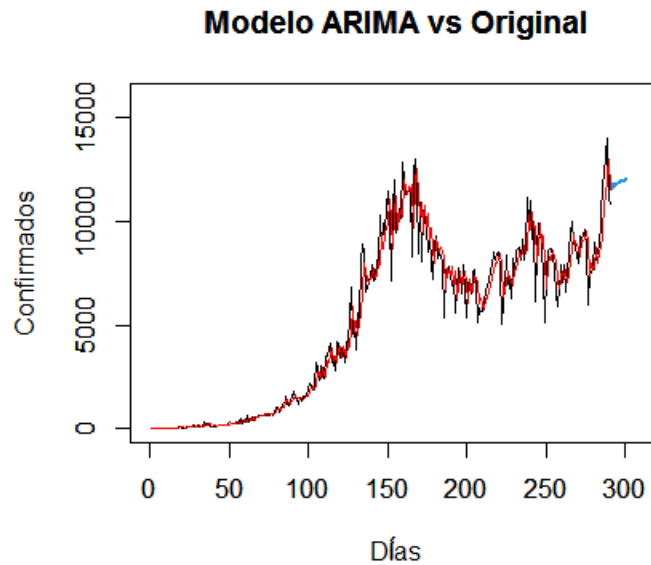


Figura 20: Pronósticos mediante el modelo ARIMA

En R existe la función *auto.arima* del paquete estadístico *forecast* el cual simplifica el análisis de la autocorrelación y la autocorrelación parcial, esta función se creó dado el proceso se realizaba al tanteo y generaba mayor gasto de tiempo en el procesamiento de los resultados. De igual manera, no siempre la función encuentra el mejor modelo y por esta razón es recomendable realizar el modelo de ambas formas.

```

Series: serie.colombia
ARIMA(0,1,3) with drift

Coefficients:
      ma1      ma2      ma3      drift
-0.4072 -0.1171 -0.0836 39.8217
s.e.    0.0592  0.0590  0.0533 22.0592

sigma^2 estimated as 915904:  log likelihood=-2400.16
AIC=4810.33  AICc=4810.54  BIC=4828.68

Training set error measures:
              ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set -0.585584 948.7713 602.0472 -Inf  Inf  0.9100531 -0.005533944

```

Figura 21: Propuesta de modelo con la función `auto.arima` del paquete `forecast` de R.

Para el caso del presente proyecto se realizó un modelo mediante `auto.arima` el cual arrojó los siguientes grados **AR(0)**, **I(1)**, **MA(3)** para obtener un **ARIMA(0,1,3)** donde el error **RMSE** es 948.77, mientras que el **RMSE** del modelo **ARIMA(1,1,3)** propuesto analizando los autocorrelogramas es 944.17, por lo que el mejor modelo es el propuesto mediante análisis.

## Comparación de modelos

Para saber cual modelo de los propuestos es mejor analizamos los errores Desviación Absoluta de la Media (DAM), Raíz del error cuadrático medio (RMSE), Error medio cuadrático (EMC), los resultados obtenidos para los 4 modelos estimados fueron:

MODELO	DAM	RMSE	EMC
Holt	638.66	1029.61	338344.65
Holt Winters	617.29	973.39	1054456.2
Media Movil	349.7	556.31	37.65
ARIMA	589.41	944.17	183993.25

Cuadro 9: Errores de los modelos

Como podemos ver, el modelo que mejor aproxima es el de medias móviles para  $n = 3$  dado que tiene los menores errores respecto a los modelos de Holt-Winters, Holt y ARIMA.

## Pronosticos vs Datos reales

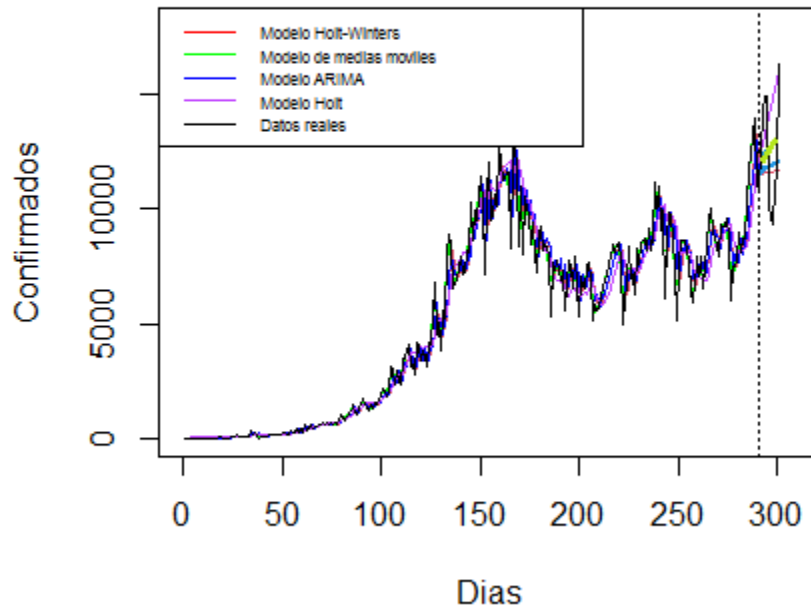


Figura 22: Datos reales vs modelos

En la imagen (22) podemos apreciar una comparación entre los datos reales y los pronósticos dados por nuestros modelos, como podemos ver en términos generales las estimaciones de todos los modelos son muy similares a los datos reales.

Fecha	ARIMA	Medias Móviles	Holt	Holt-Winters	Datos Reales
22-dic,	11532.39	12119.95	12447.21	11527.72	12526
23-dic.	11843.00	12227.86	12839.93	11547.03	14234
24-dic.	11692.96	12335.76	13232.64	11566.33	14940
25-dic.	11854.31	12443.66	13625.36	11585.64	14941
26-dic.	11822.29	12551.57	14018.08	11604.94	10196
27-dic.	11904.44	12659.47	14410.80	11624.25	9594
28-dic.	11921.41	12767.38	14803.52	11643.55	9310
29-dic.	11974.71	12875.28	15196.24	11662.86	11015
30-dic.	12008.11	12983.18	15588.96	11682.16	11639
31-dic.	12052.26	13901.09	15981.67	11701.47	16314

Cuadro 10: Comparación de pronósticos

Por otro lado, es importante comparar el estudio realizado en el presente proyecto de grado con trabajos realizados por expertos en el tema como lo es el Observatorio de

Sistemas de Salud SALUTIA <sup>4</sup>, por esta razón revisamos el pronóstico realizado por esta entidad para comparar. Esta entidad implementó el modelo SIR, el cual estima una tasa de infección y una tasa de recuperación, los supuestos que tuvieron en cuenta fueron los siguientes:

- La población está uniformemente distribuida y su tamaño no varía.
- No hay nacimientos ni muertes.
- Tasas constantes de infección y recuperación.
- Los infectados están uniformemente distribuidos en la población.

Los parámetros del modelo se obtienen minimizando la función objetivo que es la suma de los cuadrados para residuos de valores y la suma de los cuadrados de los residuos. El error que se maneja en el estudio es de  $\pm 3 * RMSE$ . La gráfica obtenida por ellos es la siguiente:

**Curva de pronóstico de contagios diarios  
(+ contagios diarios reportados)**

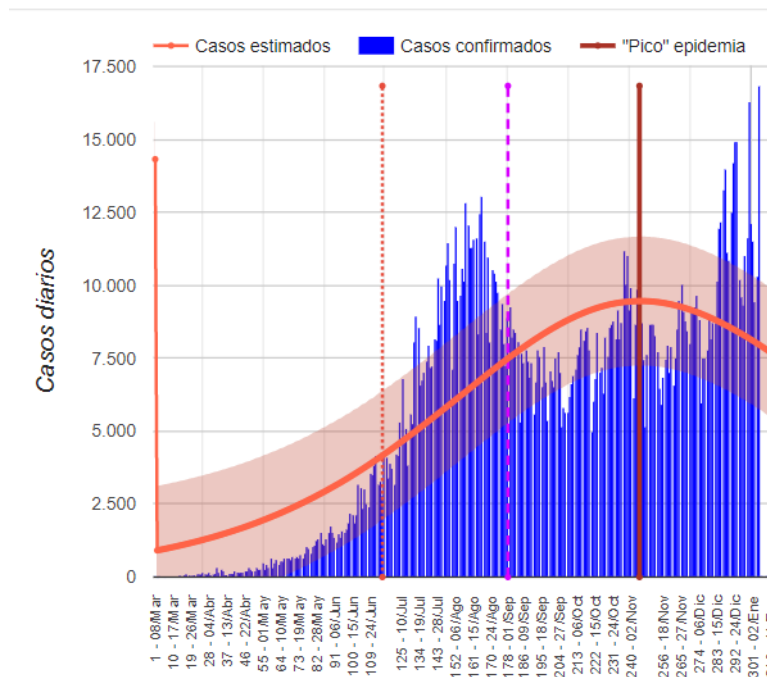


Figura 23: Pronóstico realizado por SALUTIA. Tomado de: Organización Salutia (24)

<sup>4</sup>La fundación salutia es una organización sin ánimo de lucro que nace en el año 2007 desarrolla sus actividades como centro de estudios, innovación e investigación en el sector salud, en el marco del sistema de ciencia, tecnología e innovación colombiano (24).

Podemos ver que el pronóstico de SALUTIA de casos diarios actualizado al día 13 de enero de 2021, después del 20 de octubre empieza a decrecer y para el día 31 de diciembre los casos confirmados serían aproximadamente 7100, este modelo no es totalmente acertado dado que para los últimos días de diciembre según el modelo los confirmados estarían por debajo de los 10000 casos mientras que con los datos reales del Instituto Nacional de Salud vemos que los casos han llegado casi a 15000 diarios. Esto se debe dado que el modelo SIR calcula los parámetros de infección y recuperación y basado en esto pronóstica los valores futuros.

En la fundación SALUTIA presentan diariamente una actualización de sus modelos y casos confirmados diarios que permiten saber que tan eficiente ha sido el modelo propuesto por ellos. Como se presencia en la gráfica al principio de la pandemia el modelo pronosticaba de manera eficiente pero con el pasar de los días y por el no cuidado de los ciudadanos últimamente se ha presentado un incremento en los casos diarios de contagios.

Los modelos estimados en este estudio dependen de los datos anteriores vemos que el pronóstico se mantiene creciente muy similar a los datos reales.

## Dashboard

Para poder mostrar a las personas interesadas la cantidad de contagiados, muertos y recuperados hasta el día 31 de diciembre se realizó un dashboard dinámico con la ayuda del lenguaje estadístico R. Para esto se utilizó un archivo espacial que tiene la información territorial de Colombia y con los datos tomados del Instituto Nacional de Salud se muestran la cantidad de personas que han sido afectadas por el COVID-19 por cada departamento como se muestra en la figura (24).

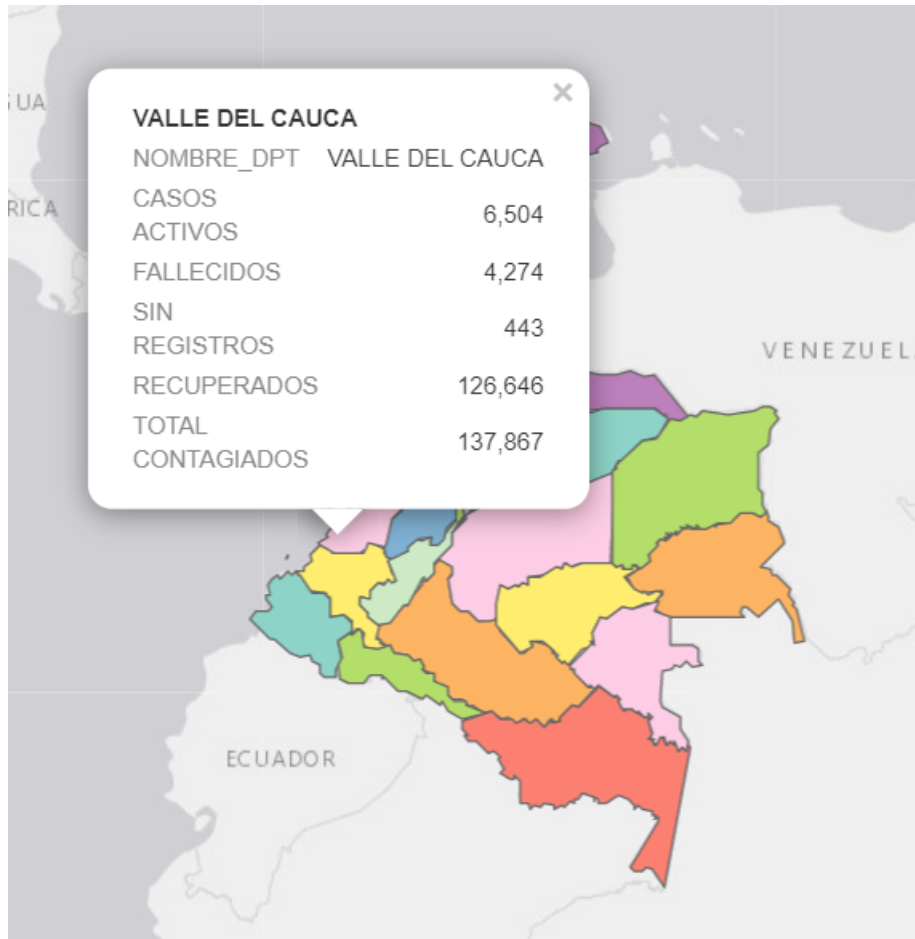


Figura 24: Dashboard de Colombia, 31 de diciembre de 2020. <https://nativanegas.github.io/index.html>

## 4. CONCLUSIONES

En el presente trabajo de investigación se realizó un estudio previo acerca de modelos de predicción estadísticos y matemáticos con el fin de pronosticar la cantidad de contagiados diarios de COVID-19 que se iba a presentar en Colombia para el día 31 de diciembre de 2020, adicionalmente se realizaron lecturas acerca de series de tiempo dado que se trabajó con la serie de tiempo dada por los confirmados diarios tomada de la base de datos del Instituto Nacional de Salud.

Se trabajaron finalmente los modelos *Holt-Winters*, *Holt*, *ARIMA* y *Medias Móviles* dado que son cuatro de los modelos que mejor trabajan con series de tiempo, para realizar las estimaciones y pronósticos utilizamos el lenguaje estadístico R el cual tiene librerías avanzadas para la construcción de este tipo de modelos. Para saber cuál modelo propuesto es el mejor se analizaron tres tipos de errores los cuales fueron, Desviación Absoluta de la Media (DAM), Error Medio Cuadrático (EMC) y Raíz del Error Cuadrático Medio (RMSE), los resultados dados por estos errores mostraron que el mejor modelo para pronosticar esta serie fue el de medias móviles con  $n = 3$  seguido del modelo *ARIMA*(1, 1, 3).

Por último, se plantea un dashboard que muestra la cantidad de contagiados, muertos y recuperados por departamento, este se realiza con el fin de dar a conocer al pueblo colombiano de manera fácil la repercusión del COVID-19 en nuestro país.

Para próximas investigaciones se deja abierta la oportunidad de trabajar con diversos métodos estadísticos-matemáticos aplicados a series de tiempo con el fin de conocer el impacto de la vacuna en Colombia que se empezará a suministrar en el año 2021.

## 5. ANEXOS

A continuación se presentarán los códigos utilizados para el modelamiento del problema:

### 5.1. Código principal

```
1 library(tidyverse)
2 library(RSocrata)
3 library(dygraphs)
4 library(xts)
5 library(incidence)
6 library(aTSA)
7 library(lmtest)
8 library(forecast)
9 library(dplyr)
10 library(seastests)
11 library(trend)
12 library(xtable)
13 library(graphics)
14 library(readxl) #Importa archivos de excel a R.
15 token <- "HufIk4NUOvHkCccqlJjKcJVqm" #Es una herramienta para poder ...
      descargar los datos desde el INS.
16 #Descarga los datos desde la pagina de la INS.
17 df.ins=read.csv("/Users/Natalia/Desktop/TESIS/2020-2/DOCUMENTOS ...
      FINAL TESIS/21dic.csv")
18 df.ins$fecha_inicio_sintomas <- ...
      as.Date(df.ins$fecha_inicio_sintomas,tryFormats = c("%d-%m-%Y", ...
      "%d/ %m/ %Y"))
19 df.ins$fecha_de_notificaci_n <- ...
      as.Date(df.ins$fecha_de_notificaci_n, tryFormats = ...
      c("%d-%m-%Y", "%d/ %m/ %Y"))
20 df.ins$fecha_muerte <- as.Date(df.ins$fecha_muerte, tryFormats = ...
      c("%d-%m-%Y", "%d/ %m/ %Y"))
21 df.ins$fecha_recuperado <- as.Date(df.ins$fecha_recuperado, ...
      tryFormats = c("%d-%m-%Y", "%d/ %m/ %Y"))
22 df.ins$fecha_reporte_web <- as.Date(df.ins$fecha_reporte_web, ...
      tryFormats = c("%d-%m-%Y", "%d/ %m/ %Y"))
23 df.ins$fecha_diagnostico <- as.Date(df.ins$fecha_diagnostico, ...
      tryFormats = c("%d-%m-%Y", "%d/ %m/ %Y"))
24 #De la 11 a la 16 convierte en formato fecha corta.
25
26 df.ins$confirmados <- "Confirmados" #Crea una columna para tener un ...
      conteo de los confirmados diarios.
27
28 df.ins <- df.ins %>%
29   dplyr::select(id_de_caso,
30                 ciudad_municipio_nom,
31                 fecha_inicio_sintomas,
32                 confirmados,
```

```

33         fecha_de_notificaci_n,
34         fecha_diagnostico,
35         fecha_muerte,
36         fecha_reporte_web,
37         everything()) #Reorganiza la base.
38
39 #####CONFIRMADOS#####
40 inc.casos.colombia.confirmados <- incidence(df.ins$fecha_reporte_web,
41     groups = df.ins$confirmados) #Hace un conteo de cantidad de
42 #confirmados por fecha.
43 serie.colombia <- xts(x = inc.casos.colombia.confirmados$counts,
44     order.by = ...
45     inc.casos.colombia.confirmados$dates) ...
46     #Crea una serie de tiempo con la
47 #cantidad de confirmados.
48
49 plot1_time <- dygraph(serie.colombia) %>%
50     dyRangeSelector(dateWindow = c("2020-03-03", "2020-08-08")) ...
51     #Grafica la serie de tiempo.
52
53 plot1_time
54 ts_colombia<- ts(serie.colombia) #Convierte en una serie de tiempo ...
55     tipo ts la serie anterior.
56
57 plot(ts_colombia,main="Curva de contagios en Colombia a o ...
58     2020",xlab="Dias",ylim=c(0,16000),xlim=c(0,310))
59
60 abline(v=c(18,31,60,105,144,165,177,187,270,296),col=c("pink","red"
61     ,"green","black","brown","darkgoldenrod","blue",
62     "darkturquoise","deeppink","darkorchid4"))
63
64 text(18,15500,"1",col="black",adj=c(0,-.1))
65 text(31,15500,"2",col="black",adj=c(0,-.1))
66 text(60,15500,"3",col="black",adj=c(0,-.1))
67 text(105,15500,"4",col="black",adj=c(0,-.1))
68 text(144,15500,"5",col="black",adj=c(0,-.1))
69 text(165,15500,"6",col="black",adj=c(0,-.1))
70 text(177,15500,"7",col="black",adj=c(0,-.1))
71 text(187,15500,"8",col="black",adj=c(0,-.1))
72 text(270,15500,"9",col="black",adj=c(0,-.1))
73 text(296,15500,"10",col="black",adj=c(0,-.1))
74
75 #####PRUEBAS ESTACIONALIDAD#####
76 #H0: La serie no tiene estacionalidad.
77 #H1: La serie tiene etacionalidad
78 #Si valor-p es menor a 0.05 se rechaza H0.
79 qs(ts_colombia) #Prueba para verificar la estacionalidad de una ...
80     serie, como el valor-p es mayor a 0.05
81 #entonces se dice que la serie no es estacional
82 isSeasonal(ts_colombia) #Dice si la serie es estacional
83
84 #####PRUEBA TENDENCIA#####
85 #H0: La serie no tiene tendencia.
86 #H1: La serie tiene tendencia.
87 #Si valor-p es menor a 0.05 se rechaza H0.
88 cs.test(ts_colombia) #Cox and stuart test, se usa para saber si una ...
89     serie tiene tendencia.
90
91 #El resultado es que tiene tendencia monotona.

```

```

78
79 #####MODELOS#####
80 #####Holt.Winters#####
81 #Se toma el caso multiplicativo por que la serie tiene tendencia
82 modHW=HoltWinters(ts_colombia,seasonal="multiplicative",gamma=FALSE)
83 predmodHW=predict(modHW,n.ahead=10) #NA INTERVAOS DE PREDICCION ...
      HELP MEEEEEEEEEEEEEEEEEE
84 predmodHW
85 xtable(predmodHW,caption="Pronostico modelo Holt-Winters")
86 plot(modHW,predmodHW,ylab="Confirmados",xlab="Das",main="Modelo ...
      Holt-Winters vs Original")
87 DAM_HW= sum(abs(ts_colombia[-c(1:2),]-
88             modHW$fitted[,1]))/length(ts_colombia[-c(1:2),])
89 RMSE_HW = sqrt(sum((modHW$fitted[,1]
90                  -ts_colombia[-c(1:2),])^2)/length(ts_colombia[-c(1:2),]))
91 EMC_HW=((sum(ts_colombia[-c(1:2),] - ...
          modHW$fitted[,1]))^2)/length(ts_colombia[-c(1:2),])
92 #####Holt#####
93 modhol=Holt(ts_colombia,plot=FALSE)
94 holt_prediccion=predict(modhol$estimate)
95 holt_prediccion=data.frame(holt_prediccion)
96 holt_prediccion[1]
97 xtable(holt_prediccion, caption="Pronostico modelo Holt")
98 modhol_esti=data.frame(modhol$estimate)
99 write.csv(holt_prediccion[1],"holt_pred.csv")
100 write.csv(modhol_esti,"modholt_es.csv")
101 holt_pred=read.csv("/Users/Natalia/Desktop/TESIS/2020-2/DOCUMENTOS ...
      FINAL TESIS/modholt_es.csv")
102 holt_pred=ts(holt_pred)
103 plot(ts_colombia,axes=FALSE,xlab=" ", ylab=" ...
      ",ylim=c(0,17000),xlim=c(0,310))
104 par(new=TRUE)
105 plot(holt_pred,xlab="Dias",ylab="Confirmados",main="Modelo Holt vs ...
      Original",col="red",
106       ylim=c(0,17000),xlim=c(0,310))
107 par(new=TRUE)
108 DAM_H= sum(abs(ts_colombia-modhol$estimate))/length(ts_colombia)
109 RMSE_H = sqrt(sum((modhol$estimate-ts_colombia)^2)/length(ts_colombia))
110 EMC_H=((sum(ts_colombia - modhol$estimate))^2)/length(ts_colombia)
111 #####Medias ...
      Moviles#####
112 #mejor
113 media3=ma(ts_colombia,3)
114 media3 #correr para comparar
115 mediamov3=na.omit(media3)
116 pred=predict(mediamov3,level=FALSE)
117 pred
118 plot(ts_colombia,axes=FALSE,xlab=" ",ylab=" ...
      ",ylim=c(0,17000),xlim=c(0,310))
119 par(new=TRUE)
120 plot(pred,ylim=c(0,17000),xlim=c(0,310),main="Modelo de medias ...
      mviles vs Original",xlab="Das",ylab="Confirmados",col="red")

```

```

121 DAM_MA3= sum(abs(ts_colombia-mediakov3))/length(ts_colombia)
122 RMSE_MA3 = sqrt(sum((mediakov3-ts_colombia)^2)/length(ts_colombia))
123 EMC_MA3=((sum(ts_colombia - mediakov3))^2)/length(ts_colombia)
124 pred=as.data.frame(pred)
125 xtable(pred,caption="Pronostico de medias moviles")
126 #####Medias moviles ...
127     7#####
128 media7=ma(ts_colombia,7)
129 media7 #correr para comparar
130 mediakov7=na.omit(media7)
131 pred=predict(mediakov7,level=0.80)
132 plot(ts_colombia,axes=FALSE,xlab=" ")
133 par(new=TRUE)
134 plot(pred,col = "red",main="Medias Moviles",xlab="Dias")
135 DAM_MA7= sum(abs(ts_colombia-mediakov7))/length(ts_colombia)
136 RMSE_MA7 = sqrt(sum((mediakov7-ts_colombia)^2)/length(ts_colombia))
137 EMC_MA7=((sum(ts_colombia - mediakov7))^2)/length(ts_colombia)
138 #####Medias moviles ...
139     9#####
140 media9=ma(ts_colombia,9)
141 media9 #correr para comparar
142 mediakov9=na.omit(media9)
143 pred=predict(mediakov9,level=0.80)
144 plot(ts_colombia,axes=FALSE,xlab=" ")
145 par(new=TRUE)
146 plot(pred,col = "red",main="Medias Moviles",xlab="Dias")
147 DAM_MA9= sum(abs(ts_colombia-mediakov9))/length(ts_colombia)
148 RMSE_MA9 = sqrt(sum((mediakov9-ts_colombia)^2)/length(ts_colombia))
149 EMC_MA9=((sum(ts_colombia - mediakov9))^2)/length(ts_colombia)
150 MEDIAMOV=c("3","7","9")
151 DAMMA=c(DAM_MA3,DAM_MA7,DAM_MA9)
152 RMSEMA=c(RMSE_MA3,RMSE_MA7,RMSE_MA9)
153 EMCMA=c(EMC_MA3,EMC_MA7,EMC_MA9)
154 data.frame(cbind(MEDIAMOV,DAMMA,RMSEMA,EMCMA))
155 xtable(data.frame(cbind(MEDIAMOV,DAMMA,RMSEMA,EMCMA)),
156     caption="Errores de medias moviles")
157 #####
158 #Antes de correr esto corre el cdigo ARIMA.
159 MODELO=c("Holt","Holt Winters","Media Movil","ARIMA")
160 DAM=c(DAM_H,DAM_HW,DAM_MA3,DAM_ARI)
161 RMSE=c(RMSE_H,RMSE_HW,RMSE_MA3,RMSE_ARI)
162 EMC=c(EMC_H,EMC_HW,EMC_MA3,EMC_ARI)
163 data.frame(cbind(MODELO,DAM,RMSE,EMC))
164 xtable(data.frame(cbind(MODELO,DAM,RMSE,EMC)),caption="Errores de ...
165     los modelos")
166 #####
167 df.ins31=read.csv("/Users/Natalia/Desktop/TESIS/2020-2/DOCUMENTOS ...
168     FINAL TESIS/31dic.csv")
169 df.ins31=filter(df.ins31,fecha_reporte_web<="2020-12-31")
170 df.ins31$confirmados <- "Confirmados" #Crea una columna para tener ...
171     un conteo de los confirmados diarios.
172 df.ins31 <- df.ins31 %>%

```

```

168 dplyr::select(id_de_caso,
169               ciudad_municipio_nom,
170               fecha_inicio_sintomas,
171               confirmados,
172               fecha_de_notificaci_n,
173               fecha_diagnostico,
174               fecha_muerte,
175               fecha_reporte_web,
176               everything()) #Reorganiza la base.
177 inc.casos.colombia.confirmados31 <- ...
178   incidence(df.ins31$fecha_reporte_web,
179   groups = df.ins31$confirmados) #Hace un conteo de cantidad de
180   #confirmados por fecha.
181   serie.colombia31 <- xts(x = inc.casos.colombia.confirmados31$counts,
182   order.by = ...
183   inc.casos.colombia.confirmados31$dates)
184   ts_colombia31<- ts(serie.colombia31) #Convierte en una serie de ...
185   tiempo tipo ts la serie anterior.
186
187   pred=predict(mediamov3, level=FALSE)
188   forecast1=forecast(arima, level=FALSE)
189   holt_prediccion=predict(modhol$estimate)
190
191   plot(modHW, predmodHW, main=" ", xlab=" ", ylab=" ...
192   ", col="white", axes=FALSE, ylim=c(0, 18000), xlim=c(0, 310)) #HW
193   par(new=TRUE)
194   plot(pred, ylim=c(0, 18000), xlim=c(0, 310), main=" ", xlab=" ", ylab=" ...
195   ", col="green", axes=FALSE) #MM
196   par(new=TRUE)
197   plot(forecast1, main=" ", xlab=" ", ylab=" ...
198   ", col="black", ylim=c(0, 18000), xlim=c(0, 310), axes=FALSE) #ARIMA
199   par(new=TRUE)
200   plot(forecast1$fitted, main=" ", xlab=" ", ylab=" ...
201   ", col="blue", ylim=c(0, 18000), xlim=c(0, 310), axes=FALSE) #ARIMA
202   par(new=TRUE)
203   plot(holt_pred, xlab=" ", ylab=" ", main=" ", col="darkorchid1",
204   ylim=c(0, 18000), xlim=c(0, 310), axes=FALSE)
205   par(new=TRUE)
206   plot(ts_colombia31, ylab="Confirmados", xlab="Dias", ylim=c(0, 18000),
207   xlim=c(0, 310), col="black",
208   main="Pronosticos vs Datos reales")
209   labs=c("Modelo Holt-Winters", "Modelo de medias moviles", "Modelo ...
210   ARIMA", "Modelo Holt", "Datos reales")
211   legend("topleft", lty = rep(1, 5), col = ...
212   c("red", "green", "blue", "darkorchid1", "black"), legend = ...
213   labs, cex=0.5)

```

## 5.2. Modelo ARIMA

```
1 #Correr librerias y lineas 14 a 36 del codigo_tesis
2
3 inc.casos.colombia.confirmados <- incidence(df.ins$fecha_reporte_web,
4     groups = df.ins$confirmados) #Hace un conteo de cantidad
5 #de confirmados por fecha.
6 serie.colombia <- xts(x = inc.casos.colombia.confirmados$counts,
7     order.by = inc.casos.colombia.confirmados$dates) #Crea una serie de
8 #tiempo con la cantidad de confirmados.
9
10 ts_colombia<- ts(serie.colombia) #Convierte en una serie de tiempo ...
11     tipo ts la serie anterior.
12
11 plot(ts_colombia)
13 plot(acf(ts_colombia)) #autocorrelacion
14 #desco=decompose(ts_colombia)
15 ndiffs(ts_colombia) #Tendencia
16 nsdiffs(ts_colombia) #No seasonal
17 #Cuando el modelo tiene tendencia se hace diff para eliminarla.
18 #Cuando en un modelo se hace diff pasa a ser ARIMA. ARIMA: Proceso ...
19     Auto regresivo integrado con medias moviles
19 diff.col=diff(ts_colombia)
20 ndiffs(diff.col)
21 plot(diff.col)
22 plot(acf(diff.col)) #Para modelo AR #autocovarianza. Si la ...
23     autocovarianza se va a 0 quiere decir que la
24 #serie ya no tiene tendencia y podemos seguir trabajando con ella.
25 plot(pacf(diff.col)) #Para modelo MA #autocovarianza parcial o ...
26     autocorrelacion/// autocovarianza(h)/autocovarianza(0)
27
28 arima=Arima(ts_colombia,order = c(1,1,3),seasonal = ...
29     list(order=c(1,1,1),period=1))
30
30 forecast1=forecast(arima,level=FALSE)
31 forecast1$fitted #Datos pronosticados por R
32 tabla_arima=forecast1$mean
33 xtable(tabla_arima,caption="Modelo ARIMA")
34 plot(forecast1,main="Modelo ARIMA vs Original",xlab=" ",ylab=" ...
35     ",col="black",ylim=c(0,16000),xlim=c(0,310))
36 par(new=TRUE)
37 plot(forecast1$fitted,main="Modelo ARIMA vs ...
38     Original",xlab="Das",ylab="Confirmados",col="red"
39     ,ylim=c(0,16000),xlim=c(0,310))
40 DAM_ARI= round(sum(abs(ts_colombia
41     -forecast1$fitted))/length(ts_colombia),2)
42 RMSE_ARI = round(sqrt(sum((forecast1$fitted
43     -ts_colombia)^2)/length(ts_colombia)),2)
```

```

42 EMC_ARI=round(((sum(ts_colombia - ...
      forecast1$fitted))^2)/length(ts_colombia),2)
43 AIC(arima)
44
45
46 ##### AUTO-ARIMA #####
47 ajuste=auto.arima(y=serie.colombia)
48
49 summary(ajuste)
50 ## predicciones
51 predicciones=forecast(ajuste, h=15)

```

### 5.3. Gráficos

```

1 library(reshape2)
2 library(ggplot2)
3 library(plyr)
4
5 df.ins=read.csv("/Users/Natalia/Desktop/TESIS/2020-2/DOCUMENTOS ...
      FINAL TESIS/31dic.csv")
6 df.ins$fecha_inicio_sintomas <- ...
      as.Date(df.ins$fecha_inicio_sintomas,tryFormats = c("%d-%m-%Y", ...
      "%d/%m/%Y"))
7 df.ins$fecha_de_notificaci_n <- ...
      as.Date(df.ins$fecha_de_notificaci_n, tryFormats = ...
      c("%d-%m-%Y", "%d/%m/%Y"))
8 df.ins$fecha_muerte <- as.Date(df.ins$fecha_muerte, tryFormats = ...
      c("%d-%m-%Y", "%d/%m/%Y"))
9 df.ins$fecha_recuperado <- as.Date(df.ins$fecha_recuperado, ...
      tryFormats = c("%d-%m-%Y", "%d/%m/%Y"))
10 df.ins$fecha_reporte_web <- as.Date(df.ins$fecha_reporte_web, ...
      tryFormats = c("%d-%m-%Y", "%d/%m/%Y"))
11 df.ins$fecha_diagnostico <- as.Date(df.ins$fecha_diagnostico, ...
      tryFormats = c("%d-%m-%Y", "%d/%m/%Y"))
12 #De la 11 a la 16 convierte en formato fecha corta.
13
14 df.ins$confirmados <- "Confirmados" #Crea una columna para tener un ...
      conteo de los confirmados diarios.
15
16 df.ins <- df.ins %>%
17   dplyr::select(id_de_caso,
18     ciudad_municipio_nom,
19     fecha_inicio_sintomas,
20     confirmados,
21     fecha_de_notificaci_n,
22     fecha_diagnostico,
23     fecha_muerte,
24     fecha_reporte_web,
25     everything()) #Reorganiza la base.

```

```

26
27 #####UNIVARIADO#####
28 #####ATENCIoN#####
29 df.ins$ubicacion[df.ins$ubicacion=="N/A"]="N/A"
30 df.ins$ubicacion=tolower(df.ins$ubicacion)
31 ubicacion=table(df.ins$ubicacion) #tabla de pacientes dependiendo ...
    la atencion que reciben
32 recuperado=table(df.ins$recuperado)
33 recuperado[5]
34 atencion=c(ubicacion,recuperado[5])
35 etiquetas_atencion=c("Casa", "Fallecido", "Hospital", "UCI", ...
    "N/A", "Recuperado")
36 datas_atencion=c(atencion[1],atencion[2],atencion[3]
37     ,atencion[4],atencion[5],atencion[6])
38 datos_atencion=data.frame(etiquetas_atencion,datas_atencion)
39 colores=c("darkolivegreen1", "darkorchid1", "pink", "orange"
40     , "grey", "darksalmon", "lightblue")
41 porcentajes_atencion = ...
    as.numeric(round(((prop.table(datos_atencion$datas_atencion))
42     *100),2))
43 porcentajes_atencion=paste(porcentajes_atencion, "%", sep="")
44
45
46 colores=c("darkolivegreen1", "darkorchid1", "pink", "orange", "grey"
47     , "darksalmon", "lightblue")
48 diag_bar_atencion=barplot(atencion,col=colores,ylim = ...
    c(0,1700000),main="Atencin a contagiados por COVID-19, ...
    Colombia 2020")
49 text(cex=1, x=diag_bar_atencion, y=-5, xpd=TRUE, srt=0, pos=1)
50 text(diag_bar_atencion, datas_atencion, labels = ...
    porcentajes_atencion,pos=3)
51
52 #####SEXO#####
53 df.ins$sexo=tolower(df.ins$sexo)
54 sexo=table(df.ins$sexo) #tabla de contagiados por sexo
55 #####TIPO#####
56 df.ins$fuente_tipo_contagio=tolower(df.ins$fuente_tipo_contagio)
57 tipo=table(df.ins$fuente_tipo_contagio)
58 tipo[1]=as.numeric(tipo[1])
59 tipo[2]=as.numeric(tipo[2])
60 tipo[3]=as.numeric(tipo[3])
61
62 etiquetas=c("[0,15]", "(15,30]", "(30,45]", "(45,60]", "(60,75]"
63     , "(75,90]", "(90,+)" )
64 datas_tipo=c(tipo[1],tipo[2],tipo[3])
65 datos_fall=data.frame(etiquetas,datas_fall)
66 colores=c("darkolivegreen1", "darkorchid1", "pink", "orange", "grey",
67     "darksalmon", "lightblue")
68 porcentajes_tipo = as.numeric(round(((prop.table(datos_tipo))*100),2))
69 porcentajes_tipo=paste(porcentajes_tipo, sep="")
70

```

```

71 diag_bar_tipo=barplot(datas_tipo,col=cbind("lightblue","pink"),ylim ...
    = c(0,1600000))
72 text(diag_bar_tipo, datas_tipo, labels = porcentajes_tipo,pos=3)
73
74 etiquetas_tipo=c("En estudio", "Importado","Relacionado")
75 etiquetas_tipo=paste(etiquetas_tipo,porcentajes_tipo)
76 etiquetas_tipo=paste(etiquetas_tipo,"%",sep="")
77 colors=c("lightblue","red","pink")
78 pie(tipo,etiquetas_tipo,main="Tipo de contagio por COVID-19, ...
    Colombia 2020",col=colors) #Grafica la tabla anterior.
79
80
81 #####EDAD#####
82 df.ins$edad=as.numeric(df.ins$edad) #Convierte a un valor numerico ...
    la edad de los pacientes
83
84 #porcentajes <- ...
    as.numeric(round(((prop.table(table(df.ins$edad)))*100),2))
85 tablaedad=table(df.ins$edad)
86 tablaedad=data.frame(tablaedad)
87 names(tablaedad)=c("Edades", "Contagiados")
88 tablaedad$Edades=as.numeric(tablaedad$Edades)
89 tablaedad$Contagiados=as.numeric(tablaedad$Contagiados)
90
91 tablaedad1=sum(tablaedad$Contagiados[which(tablaedad$Edades <= 15)])
92 tablaedad2=sum(tablaedad$Contagiados[which(tablaedad$Edades>15 & ...
    tablaedad$Edades<=30)])
93 tablaedad3=sum(tablaedad$Contagiados[which(tablaedad$Edades>30 & ...
    tablaedad$Edades<=45)])
94 tablaedad4=sum(tablaedad$Contagiados[which(tablaedad$Edades>45 & ...
    tablaedad$Edades<=60)])
95 tablaedad5=sum(tablaedad$Contagiados[which(tablaedad$Edades>60 & ...
    tablaedad$Edades<=75)])
96 tablaedad6=sum(tablaedad$Contagiados[which(tablaedad$Edades>75 & ...
    tablaedad$Edades<=90)])
97 tablaedad7=sum(tablaedad$Contagiados[which(tablaedad$Edades>90)])
98 etiquetas=c("[0,15]", "(15,30]", "(30,45]", "(45,60]", "(60,75]"
99     , "(75,90]", "(90,+)" )
100 datas_tablaedad=c(tablaedad1,tablaedad2,tablaedad3,
101     tablaedad4,tablaedad5,tablaedad6,tablaedad7)
102 datos_tablaedad=data.frame(etiquetas,datas_tablaedad)
103 colores=c("darkolivegreen1", "darkorchid1", "pink", "orange", "grey"
104     , "darksalmon", "lightblue")
105 porcentajes_tablaedad = ...
    as.numeric(round(((prop.table(datos_tablaedad$datas_tablaedad))
106     *100),2))
107 porcentajes_tablaedad=paste(porcentajes_tablaedad,"%",sep="")
108
109 diag_bar_tablaedad=barplot(datos_tablaedad$datas_tablaedad,
110     col=colores,ylim = c(0,600000),main="Edad de contagiados ...
    por COVID-19, Colombia 2020")

```

```

111 text(cex=1, x=diag_bar_tablaedad, y=-5, etiquetas, xpd=TRUE, srt=0, ...
      pos=1)
112 text(diag_bar_tablaedad, datas_tablaedad, labels = ...
      porcentajes_tablaedad, pos=3)
113
114
115 #####DEPARTAMENTO#####
116 ###Correr codigo dash librerias y desde la linea 15 hasta la linea ...
      18 para poder correr esto###
117 df.ins$codigo=NA #Crea una nueva columna sin datos.
118 df.ins$codigo=trunc(as.numeric(df.ins$ciudad_municipio)/1000) #A la ...
      columna creada le asigna el valor del codigo divipola
119 #dividido por 1000, el cual es igual al codigo del dpto.
120 df.ins$ubicacion[df.ins$ubicacion=="N/A"]="N/A"
121 df.ins$ubicacion=tolower(df.ins$ubicacion)
122 tabla=table(df.ins$codigo,df.ins$ubicacion) #Crea una tabla que ...
      cuenta la cantidad de atendidos dependiendo del codigo de dpto.
123 tabla[,1]=tabla[,1]+tabla[,3]+tabla[,4]
124 tabla=tabla[,-c(3:4)]
125 df.ins$recuperado[df.ins$recuperado=="N/A"]="N/A"
126 df.ins$recuperado=tolower(df.ins$recuperado)
127 tabla1=table(df.ins$codigo,df.ins$recuperado)
128 tabla1=tabla1[,-c(1:3)]
129 tabla1=as.data.frame(tabla1)
130 TOTAL=tabla[,1]+tabla[,2]+tabla[,3]+tabla1
131 TOTAL=as.data.frame(TOTAL)
132 dashboard=arrange(dashboard, as.numeric(DPTO)) #Organiza de menor a ...
      mayor la base dashboard dependiendo del cod. de dpto.
133 rownames(TOTAL)=dashboard$NOMBRE_DPT
134 rownames(tabla)=dashboard$NOMBRE_DPT
135 TOTAL
136 write.csv(TOTAL, "total.csv")
137 TOTAL=read.xlsx("/Users/Natalia/Desktop/TESIS/2020-2/DOCUMENTOS ...
      FINAL TESIS/total.xlsx")
138 TOTAL=arrange(TOTAL, Contagiados)
139 xtable(TOTAL, caption="Cantidad de contagiados por departamento. ...
      Tomado de: cite{inspron}")
140 tabla
141 write.csv(tabla, "tabla.csv")
142 tabla=read.xlsx("/Users/Natalia/Desktop/TESIS/2020-2/DOCUMENTOS ...
      FINAL TESIS/tabla.xlsx")
143 tabla=arrange(tabla, Fallecidos)
144 xtable(tabla, caption="Cantidad de fallecidos por departamento. ...
      Tomado de: cite{inspron}")
145 #####BIVARIADO#####
146 #####FALLECIDOS POR EDAD#####
147 fall_edad=table(df.ins$edad,df.ins$ubicacion=="Fallecido") #Crea ...
      una tabla para saber la cantidad de fallecidos por edad.
148
149 tabla_fall_edad=data.frame(fall_edad)
150 tabla_fall_edad=filter(tabla_fall_edad, tabla_fall_edad$Freq==TRUE)
151 tabla_fall_edad=tabla_fall_edad[-2]

```

```

152 names (tabla_fall_edad)=c("Edades", "Contagiados")
153 tabla_fall_edad$Edades=as.numeric(tabla_fall_edad$Edades)
154 tabla_fall_edad$Contagiados=as.numeric(tabla_fall_edad$Contagiados)
155
156 fall1=sum(tabla_fall_edad$Contagiados[which(tabla_fall_edad$Edades ≤...
      15)])
157 fall2=sum(tabla_fall_edad$Contagiados[which(
158   tabla_fall_edad$Edades>15 & tabla_fall_edad$Edades≤30)])
159 fall3=sum(tabla_fall_edad$Contagiados[which(
160   tabla_fall_edad$Edades>30 & tabla_fall_edad$Edades≤45)])
161 fall4=sum(tabla_fall_edad$Contagiados[which(
162   tabla_fall_edad$Edades>45 & tabla_fall_edad$Edades≤60)])
163 fall5=sum(tabla_fall_edad$Contagiados[which(
164   tabla_fall_edad$Edades>60 & tabla_fall_edad$Edades≤75)])
165 fall6=sum(tabla_fall_edad$Contagiados[which(
166   tabla_fall_edad$Edades>75 & tabla_fall_edad$Edades≤90)])
167 fall7=sum(tabla_fall_edad$Contagiados[which(
168   tabla_fall_edad$Edades>90 )])
169 etiquetas=c("[0,15]", "(15,30]", "(30,45]", "(45,60]",
170             "(60,75]", "(75,90]", "(90,+)" )
171 datas_fall=c(fall1,fall2,fall3,fall4,fall5,fall6,fall7)
172 datos_fall=data.frame(etiquetas,datas_fall)
173 colores=c("darkolivegreen1", "darkorchid1", "pink", "orange",
174           "grey", "darksalmon", "lightblue")
175 porcentajes_fall = ...
      as.numeric(round(((prop.table(datos_fall$datas_fall))*100),2))
176 porcentajes_fall=paste(porcentajes_fall, "%", sep="")
177
178 diag_bar_fall=barplot(porcentajes_fall,col=colores,ylim = c(0,15000))
179 text(cex=1, x=diag_bar_fall, y=-5, etiquetas, xpd=TRUE, srt=0, pos=1)
180 text(diag_bar_fall, datas_fall, labels = porcentajes_fall,pos=3)
181
182
183 #####RECUPERADOS POR EDAD#####
184 recu_edad=table(df.ins$edad,df.ins$recuperado=="Recuperado") #Crea ...
      una tabla para saber la cantidad de recuperados por edad.
185
186 tabla_recu_edad=data.frame(recu_edad)
187 tabla_recu_edad=filter(tabla_recu_edad,tabla_recu_edad$Var2==TRUE)
188 tabla_recu_edad=tabla_recu_edad[-2]
189 names (tabla_recu_edad)=c("Edades", "Contagiados")
190 tabla_recu_edad$Edades=as.numeric(tabla_recu_edad$Edades)
191 tabla_recu_edad$Contagiados=as.numeric(tabla_recu_edad$Contagiados)
192
193 recu1=sum(tabla_recu_edad$Contagiados[which(
194   tabla_recu_edad$Edades ≤ 15)])
195 recu2=sum(tabla_recu_edad$Contagiados[which(
196   tabla_recu_edad$Edades>15 & tabla_recu_edad$Edades≤30)])
197 recu3=sum(tabla_recu_edad$Contagiados[which(
198   tabla_recu_edad$Edades>30 & tabla_recu_edad$Edades≤45)])
199 recu4=sum(tabla_recu_edad$Contagiados[which(
200   tabla_recu_edad$Edades>45 & tabla_recu_edad$Edades≤60)])

```

```

201 recu5=sum(tabla_recu_edad$Contagiados[which(
202   tabla_recu_edad$Edades>60 & tabla_recu_edad$Edades<=75)])
203 recu6=sum(tabla_recu_edad$Contagiados[which(
204   tabla_recu_edad$Edades>75 & tabla_recu_edad$Edades<=90)])
205 recu7=sum(tabla_recu_edad$Contagiados[which(
206   tabla_recu_edad$Edades>90 )])
207 etiquetas=c("[0,15]", "(15,30]", "(30,45]", "(45,60]"
208             , "(60,75]", "(75,90]", "(90,+)"
209 datas_recu=c(recu1,recu2,recu3,recu4,recu5,recu6,recu7)
210 datos_recu=data.frame(etiquetas,datas_recu)
211 colores=c("darkolivegreen1", "darkorchid1", "pink", "orange", "grey",
212           "darksalmon", "lightblue")
213 porcentajes_recu=as.numeric(round(((
214   prop.table(datos_recu$datas_recu)*100),2))
215 porcentajes_recu=paste(porcentajes_recu, "%", sep="")
216
217 diag_bar_recu=barplot(datos_recu$datas_recu,
218                       col=colores,ylim=c(0,380000))
219
220 text(cex=1, x=diag_bar_recu, y=-5, etiquetas, xpd=TRUE, srt=0, pos=1)
221 text(diag_bar_recu, datos_recu + 0.5, labels = porcentajes_recu,pos=3)
222 #####CONTAGIADOS POR EDAD Y SEXO#####
223 # https://pmoracho.github.io/blog/2019/03/10/sample-ggplot-plots/
224 df.ins$edad=as.numeric(df.ins$edad)
225 df.ins$edadR=cut(df.ins$edad, breaks = c(0,5,10,15,20,25,30,35,40,
226     45,50,55,60,65,70,75,80,85,90,95,100,125))
227 (genero=table(df.ins$edadR,df.ins$sexo))
228 elabels<-c("(0,5]", "(5,10]", ...
229            "(10,15]", "(15,20]", "(20,25]", "(25,30]", "(30,35]", "(35,40]",
230            "(40,45]", "(45,50]", "(50,55]", "(55,60]", "(60,65]", ...
231            "(65,70]", "(70,75]", "(75,80]",
232            "(80,85]", "(85,90]", "(90,95]", "(95,100]", "(100 +]")
233 d=dim(df.ins)
234 masc=round(genero[,2]/d[1]*100,2)
235 fem=round(genero[,1]/d[1]*100,2)
236
237 df=rbind.data.frame(
238   cbind.data.frame(Percentage = -masc, Group = elabels, Gender = "M"),
239   cbind.data.frame(Percentage = +fem, Group = elabels, Gender = "F")
240 )
241
242 # Make sure agelabels have the right order
243 df$Group <- factor(df$Group, levels = elabels);
244
245 # (gg) grafico de piramide
246 ggplot(data = df, aes(x = Group, y = Percentage, fill = Gender, ...
247   group = Gender)) +
248   geom_bar(data = subset(df, Gender == "F"), stat = "identity") +
249   geom_bar(data = subset(df, Gender == "M"), stat = "identity") +
250   coord_flip() +

```

```

250 geom_smooth(colour = "black", method = "loess", se = FALSE, ...
      show.legend = FALSE, size = 0.5) +
251 scale_fill_manual(values=c('#25AAE2', '#8BC540')) +
252 labs(
253   x = "Edad",
254   y = "Porcentaje",
255   title = "Edad contagiados por sexo Covid-19 Colombia 2020",
256   caption="fuente: https://www.datos.gov.co")
257 scale_y_continuous(
258   breaks = seq(-4, 4, by = 1),
259   labels = c(rev(seq(0, 4, by = 2)), seq(2, 4, by = 2)))
260 #theme_elegante()

```

## 5.4. Dashboard

```

1 library(sf) #Contiene funciones para codificar datos espaciales.
2 library(ggplot2) #Sirve para crear graficos y mapear
3 library(tmap) #Importa mapas tematicos en los cuales se visualizan ...
  las distribuciones de datos espaciales.
4 library(tmaptools) #Conjunto de herramientas para leer y procesar ...
  datos espaciales.
5 library(leaflet) #Crea y personaliza mapas interactivos usando una ...
  biblioteca de JavaScript
6 library(dplyr) #Sirve para manipular datos.
7 library(rgdal) #Proporciona enlaces a la Biblioteca de abstraccion ...
  de datos 'geoespaciales' y acceso
8 #a las operaciones de proyeccion.
9 library(sp) #Clases y metodos para datos espaciales.
10 library(readxl) #Importa archivos de excel a R.
11 library(RSocrata) #Proporciona una interaccion mas facil con los ...
  portales de datos abiertos.
12 library(mapview)
13 token <- "HufIk4NUOvHkCccqlJjKcJVqm" #Es una herramienta para poder ...
  descargar los datos desde el INS.
14 df.ins <- ...
  read.socrata("https://www.datos.gov.co/resource/gt2j-8ykr.json"
15             , app_token = token) #Descarga los datos desde la pagina de ...
  la INS.
16 df.ins=read.csv("/Users/Natalia/Desktop/TESIS/2020-2/DOCUMENTOS ...
  FINAL TESIS/31dic.csv")
17 mapa=readOGR("/Users/Natalia/Desktop/TESIS/depto/depto.shp", ...
  stringsAsFactors=FALSE) #Exporta los datos para crear el mapa ...
  de Colombia dividido por departamentos
18 dashboard=st_read("/Users/Natalia/Desktop/TESIS/depto/depto.shp"
19                  ,stringsAsFactors=FALSE) #Exporta las ...
  características de cada departamento de Colombia
20 plot(mapa,main="COLOMBIA") #Crea el mapa de Colombia
21 str(mapa) #Muestra de forma compacta la estructura interna de un ...
  objeto.

```

```

22
23 #####Correr codigo dash librerias y desde la linea 15 hasta la linea ...
    18 para poder correr esto####
24 df.ins$codigo=NA #Crea una nueva columna sin datos.
25 df.ins$codigo=trunc(as.numeric(df.ins$ciudad_municipio)/1000) #A la ...
    columna creada le asigna el valor del codigo divipola
26 #dividido por 1000, el cual es igual al codigo del dpto.
27 df.ins$recuperado[df.ins$recuperado=="N/A"]="N/A"
28 df.ins$recuperado=tolower(df.ins$recuperado)
29 tabla1=table(df.ins$codigo,df.ins$recuperado)
30 TOTAL=tabla1[,1]+tabla1[,2]+tabla1[,3]+tabla1[,4]
31 TOTAL=as.data.frame(TOTAL)
32 dashboard=arrange(dashboard, as.numeric(DPTO)) #Organiza de menor a ...
    mayor la base dashboard dependiendo del cod. de dpto.
33 rownames(TOTAL)=dashboard$NOMBRE_DPT
34 rownames(tabla1)=dashboard$NOMBRE_DPT
35 write.csv(tabla1, file="covid19.csv") #Guarda en un archivo .csv la ...
    tabla creada.
36 covid19=read.csv("/Users/Natalia/Desktop/TESIS/2020-2/DOCUMENTOS ...
    FINAL TESIS/covid19.csv") #Exporta el archivo creado anteriormente.
37 covid19=data.frame(covid19,TOTAL) #Concatena el archivo anterior ...
    con el total de contagiados.
38 names(covid19)=cbind("NOMBRE_DPT", "CASOS ACTIVOS", "FALLECIDOS", "SIN ...
    REGISTROS", "RECUPERADOS", "TOTAL CONTAGIADOS") #Renombra las ...
    columnas.
39 dashboard=dashboard[,-1] #Elimina la primera columna de la base ...
    dashboard.
40 mapview(covid19[,2],layer.name="Total")
41 names(dashboard)
42 mapa_datos=inner_join(dashboard,covid19)#Une las bases dashboard y ...
    covid 19.
43 mapa_datos=mapa_datos[,-c(2:4)] #Elimina las columnas 2 a 4 de la ...
    base creada anteriormente.
44 tm_shape(mapa_datos)+
45   tm_layout(title = "CASOS DE COVID-19 EN COLOMBIA. Realizado por: ...
    Natalia Vanegas")+
46   tm_polygons("MAP_COLORS") #Crea el mapa estatico de Colombia ...
    dependiendo de los datos anteriores.
47
48 test_map=tmap_last() #Guarda el mapa.
49 tmap_save(test_map,"mapa.html",add.titles = TRUE) #Crea el archivo ...
    HTML del mapa dinamico.

```

## Referencias

- [1] Hanke John & Reitsch Arthur. Pronósticos en los negocios, quinta edición, 1996.
- [2] Salazar Lopez Bryan. (2019). Regresión lineal. Pronóstico de la demanda: Mínimos cuadrados Recuperado de <https://www.ingenieriaindustrialonline.com/pronostico-de-la-demanda/regresion-lineal/>.
- [3] Ballou Ronald. Logística: Administración de la cadena de suministros, quinta edición, 2004.
- [4] Schroeder Roger. Administración de operaciones, 1992.
- [5] Ferrel O.C. & Hartline Michael. Estrategia de Marketing, quinta edición, 2014.
- [6] de Lara Haro, Alfonso. Medición y control de riesgos financieros, 2001.
- [7] Anonimo. (2019). Redes neuronales Recuperado de <https://sites.google.com/site/mayinteligenciartificial/unidad-4-redes-neuronales>
- [8] Anonimo. (2019). Método de Holt-Winters Recuperado de <http://previsioneventas.com/nivel-basico/metodo-de-holt-winters/>
- [9] Johns Hopkins University. (2020). Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE Recuperado de <https://github.com/CSSEGISandData/COVID-19>
- [10] Organización Mundial de la Salud. (2020). Preguntas y respuestas sobre la enfermedad por coronavirus (COVID-19) Recuperado de <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses>
- [11] Instituto Nacional de Salud de Colombia. (2020). Coronavirus (CoVid-2019) en Colombia. Recuperado de <https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx>
- [12] Organización Mundial de la Salud. (2020). Dashboard. Recuperado de <http://healthcybermap.org/WHOCOVID19/1>
- [13] Healthmap (2020). Dashboard. Recuperado de <https://healthmap.org/es/>
- [14] Kamel Boulos Maged, Geraghty Estella. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics, 2020.
- [15] John Hopkins University. Recuperado de: <https://gisanddata.maps.arcgis.com/apps/>, 2020.

- [16] Organización Mundial de la Salud, Recuperado de:<https://covid19.who.int/>, 2020.
- [17] HealthMap, Recuperado de:<https://www.healthmap.org/covid-19/>, 2020.
- [18] Garcia Perez Alfonso, Métodos avanzados de estadística aplicada: Técnicas avanzadas, 2005.
- [19] Cuartas Daniel, Arango-Londoño David, Guzman-Escarria Gissel, Muñoz Edgar, Caicedo Diana, Ortega Delia, Fandiño-Losada Andrés, Mena Jorge, Torres Miyerlandi, Barrera Lena & Méndez Fabián, Análisis espacio-temporal del SARS-COV-2 en Cali, Colombia, 2020.
- [20] Ortega-Lenis Delia, Arango-Londoño David, Muñoz Edgar, Cuartas Daniel, Caicedo Diana, Mena Jorge, Torres Miyerlandi & Mendez Fabián, Predicciones de un modelo SEIR para casos de COVID-19 en Cali, Colombia, 2020.
- [21] Pérez Abreu, Manuel Ramón & Gómez Tejeda, Jairo Jesús & Dieguez Guach, Ronny Alejandro, Características clínico-epidemiológicas de la COVID-19, Revista Habanera de Ciencias Médicas, 2020
- [22] Yanes Carlos Andres & Porras Tubercuía Alejandro, Modelos de proyección de COVID-19, Universidad del Norte, Barranquilla, Colombia, 2020
- [23] González Casimiro María Pilar, Análisis de series temporales: Modelos ARIMA, Universidad de País Vasco, 2009
- [24] Pronósticos curva contagios COVID-19 Colombia. (2020). Recuperado de <https://salutia.org/pronosticos-comportamiento-covid-19>
- [25] Instituto Nacional de Salud - Observatorio Nacional de Salud, Modelo de transmisión de coronavirus COVID-19 - Escenarios para Colombia, 2020
- [26] Administración de la producción y las operaciones conceptos, modelos y funcionamiento= Production and operations management, Everett, E., Adam, E., & Ebert, R. J. 1991
- [27] El método de Holt para elaborar pronósticos con tendencia, Izar Juan Manuel, Instituto tecnológico superior de Rioverde, 2019
- [28] Estimación de una tendencia determinista y un componente estacional, Administración y dirección de empresas Departamento de estadístico, Universidad Carlos III, Madrid, 2009
- [29] Introducción a series de tiempo, John Villavicencio, Gobierno de Perú, 2015
- [30] Series temporales, Juan Miguel Marín Diazaraque, Universidad Carlos III, Madrid, 2011