

Una consideración empírica preliminar del Coronavirus en Colombia

23 DE MARZO DE 2020

CORPORACIÓN CIPJUS

Creado por: John Michael Riveros Gavilanes



CORPORACIÓN

Una consideración empírica preliminar del Coronavirus en Colombia

Resumen: El presente documento busca presentar una consideración empírica desde la información recopilada por el Instituto Nacional de Salud (2020) en relación con los casos de contagio por Covid-19. Este estudio no busca pronosticar o proyectar la tasa de contaminados dado que no parte de las bases de la disciplina epidemiológica, sino que parte más que nada del análisis de información intrínseca en la base de datos del Instituto Nacional de Salud, sobre el contagio que empezó el 6 de marzo de 2020 en Colombia (Sudamérica) hasta la fecha de su realización (el 23 de marzo de 2020). La metodología parte de un análisis descriptivo donde se establecen las características principales de los infectados, análisis de correlaciones básicos y un conjunto de modelos de regresión que abarcan las variables de sexo, edad promedio de los infectados, número de infectados importadores del virus en el extranjero y número de infectados por relación indirecta. Los resultados evidencian que los contagios por día no se encuentran determinados por el sexo o la edad, pero si lo están con el número de importadores del virus y los infectados relacionados indirectamente.

Palabras Clave: Covid-19, Coronavirus, Regresión, Series de Tiempo, Correlaciones, Colombia.

Abstract: The present paper presents an empirical consideration from the data collected of the National Institute of Health in Colombia (2020) related to the confirmed cases of Covid-19. This study does not pursue to predict or forecast the growth of the contamination since it is not derived and does not have any linkage to the epidemiologic discipline. The methodology parts from a descriptive analysis of the principal characteristics of the cases, basic correlation analysis and three regression models which covers the variables of sex, average age, number of importers of the virus (derived of travels outside Colombia) and the number of cases which were non-directly related to the importers. The results evidence that contagious per day are not determined by age or sex, but there are statistically significant correlated with the number of importers of the virus and the number of non-directly infected.

Keywords: Covid-19, Coronavirus, Regression, Time Series, Correlations, Colombia.

Introducción

El 17 de noviembre de 2019 se detectó en el mundo el primer caso de Covid-19, popularmente llamado Coronavirus (Davidson, 2020) en el área comprendida de la ciudad de Wuhan en China, desde entonces ha tenido una propagación significativa a lo largo del globo. La expansión ha sido tal, que al menos 147 países han resultado afectados por la propagación de la infección viral desde entonces (Bryson Taylor, 2020), causando síntomas de neumonía y gripe parecidos a los causados por otros coronavirus como el síndrome respiratorio agudo grave (del inglés Severe Acute Respiratory Syndrome) -SARS- y el síndrome respiratorio de medio oriente (del inglés Middle East Respiratory Syndrome) -MERS-.

Aunque la letalidad del virus es diferencial al SARS y el MERS (Mizumoto & Chowell, 2020), su nivel expansión hizo que se volviera motivo de preocupación a lo largo de las naciones del mundo. De acuerdo con el Centro de Sistemas de Ciencia e Ingeniería -CSSE- de la Universidad John Hopkins, a la fecha de hoy, 23 de marzo de 2020, existen 353.692 casos confirmados de coronavirus en el planeta, donde 15.430 han muerto a causa del virus y 100.443 se han recuperado (CSSE, 2020).

La escala de contaminación a la fecha se presenta el gráfico siguiente y demuestra la intensidad de pandemia a nivel global.

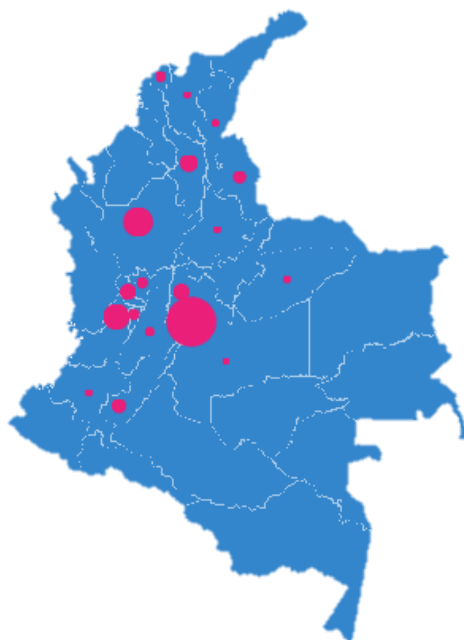
Gráfico 1 Propagación a fecha del 23 de marzo de 2020 del virus a escala mundial.



Nota: Los círculos en rojo representan por su tamaño la cantidad de casos confirmados de coronavirus en el mundo. Fuente: CSSE de la Universidad John Hopkins (2020)

En el caso colombiano, el día 6 de marzo de 2020 se confirmó el primer caso en el país, dando lugar a un cambio de los protocolos de prevención para derivar en los protocolos de contención (INS, 2020) del Covid-19.

Gráfico 2 Expansión a Nivel Colombia con fecha del 23 de marzo de 2020



Nota: Los círculos representan por su tamaño la cantidad de casos confirmados de coronavirus en el mundo. Fuente: INS (2020).

A la fecha del 23 de marzo de 2020 se destaca que existe un total de 277 casos, en los cuales Bogotá (capital del país) tiene el peso más significativo de contagios correspondiente al 40.79% de los infectados en el país, seguido por el departamento de Antioquia con el 14.4%. El grueso de la población infectada en el país se ubica en Bogotá y esto es explicado por su estrecha relación con los flujos de migrantes que llegan y salen del país.

El objetivo de este documento no es proveer un estudio epidemiológico en relación con la expansión del coronavirus, sino que busca proveer información inicial relacionada con análisis básicos de forma preliminar con enfoques de la minería de datos sobre la expansión y contagio del Covid-19 en el país, así como el comportamiento relacionado con el seguimiento del virus y su relación con variables básicas como edad, sexo y origen de la infección con la información suministrada por el Instituto Nacional de Salud.

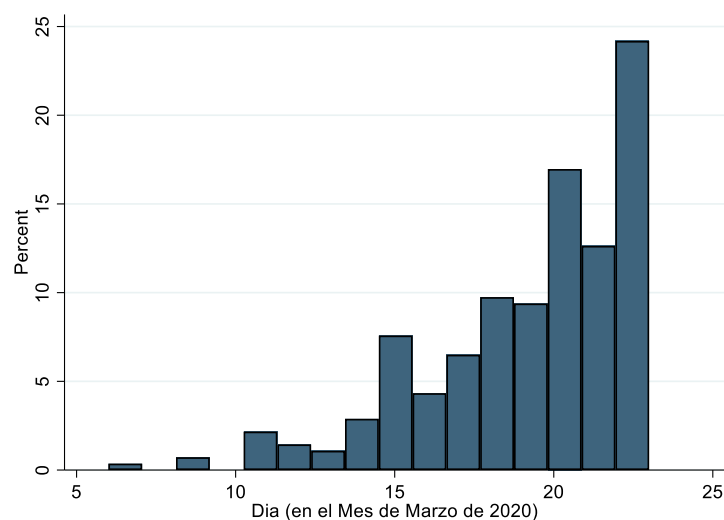
Hechos Estilizados

En esta sección se presentan las estadísticas básicas descriptivas relacionadas al número de personas contagiadas por el Covid-19 en Colombia, la información se obtiene de las series diarias publicadas por el INS (2020) desde el 6 de marzo de 2020 hasta el 23 del mismo.

Análisis de tipo corte transversal

De acuerdo con la información obtenida de las características y fechas individuales de los contagiados del Covid-19, se procede a un breve análisis de las características recopiladas en virtud de proveer información del comportamiento del contagio. En esta sección se presenta el cumulo de información que se recopiló en relación con los infectados.

Gráfico 3 Porcentaje de Infectados en Relación con tiempo y el Total desde el inicio de la infección Covid-19 Colombia



Nota: El histograma identifica de acuerdo con el total de infectados (277 en el país a marzo 23 de 2020) su proporción en relación con las fechas específicas de marzo. La información considera el número de contaminados por día en relación con el total. Fuente: Elaboración Propia con Información de INS (2020).

La progresión del número de contaminados en relación con su total es contundente, empezó su expansión el 6 de marzo, y aunque inmediatamente no se presentaron casos el día siguiente, el 8 de marzo comenzó una escalada mayor, así mismo, se puede ver un proceso secuencial de aumentos y disminuciones de la cantidad de nuevos contaminados por día en el periodo de tiempo observado, cuestión que puede dar indicios de que se configure un fenómeno cíclico/estacional de la propagación. Por el

momento, es de destacar que los principios de comportamiento en el tiempo son consistentes con las fases iniciales de una escalada mayor.

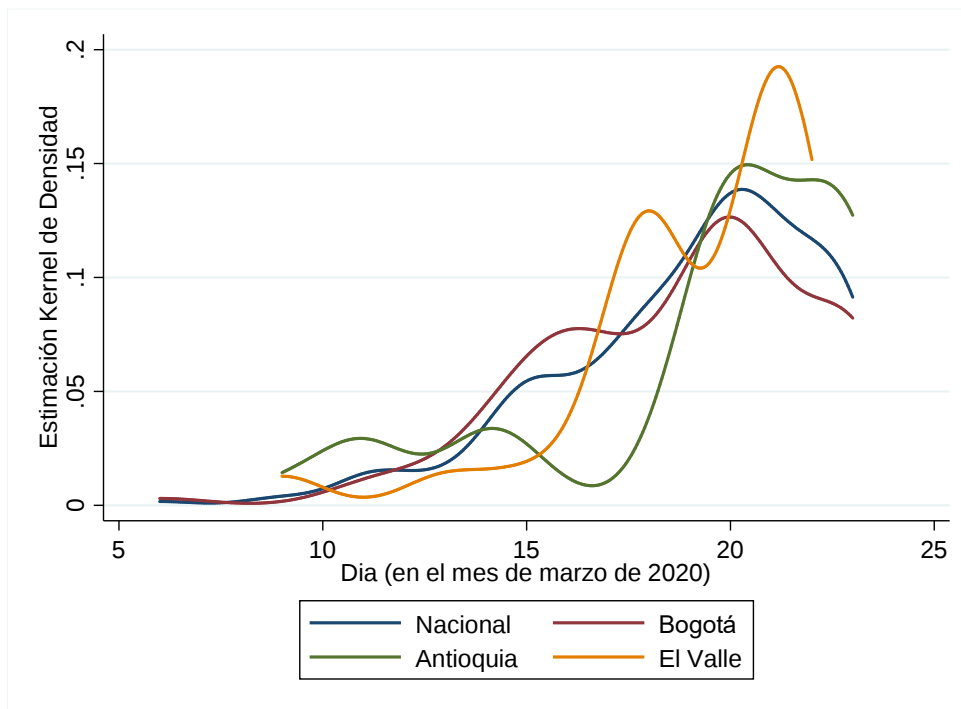
Tabla 1 Casos Confirmados por Departamento

Departamento	Freq.	Percent	Cum.
Antioquia	40	14.44	14.44
Atlántico	7	2.53	16.97
Bogotá	113	40.79	57.76
Bolívar	16	5.78	63.54
Caldas	6	2.17	65.70
Casanare	1	0.36	66.06
Cauca	2	0.72	66.79
Cesar	1	0.36	67.15
Cundinamarca	12	4.33	71.48
Huila	10	3.61	75.09
Magdalena	2	0.72	75.81
Meta	1	0.36	76.17
Norte de Santander	8	2.89	79.06
Quindío	6	2.17	81.23
Risaralda	13	4.69	85.92
San Andrés Islas	1	0.36	86.28
Santander	3	1.08	87.36
Tolima	4	1.44	88.81
Valle	31	11.19	100.00
Total	277	100.00	

Nota: Los valores de frecuencia (freq) se refiere al número de personas con Covid-19 hasta la fecha (23 de marzo de 2020). Fuente: Elaboración Propia con Información de INS (2020).

Se observa que la cantidad principal de casos confirmados la tiene Bogotá con un 40.79% del total nacional de casos, seguido por Antioquia (14.49%) y El Valle (11.19%).

Gráfico 4 Estimación Kernel de Densidad de Probabilidad Gaussiana



Nota: La estimación Kernel de densidad fue calculada para la variable de número de casos por día con la aproximación Gaussiana. Fuente: Elaboración Propia con Información de INS (2020).

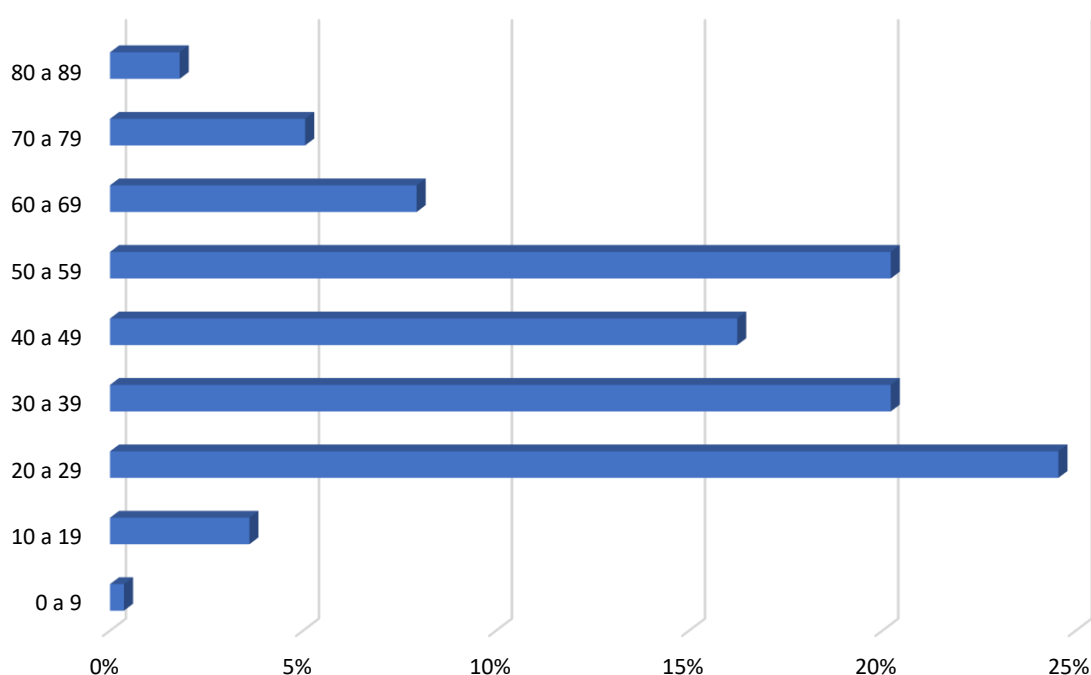
En este aspecto, se observa que la probabilidad de contagio conforme va aumentando los días va acentuándose de forma más significativa en los departamentos de El Valle, Antioquia a comparación del nivel de Bogotá y el nacional. Esto lleva a pensar que la fase de probabilidad de contagio será mayor en los departamentos en donde, se han presentado los casos más recientes. Esto se infiere por el inicio de la función diferente al punto inicial en el tiempo de Bogotá, esta brecha temporal implica un grado de

probabilidad aún mayor en el tiempo de que se propague el Covid-19, en especial en el caso de El Valle.

Es de enfatizar que la parte decreciente de las curvas no se debe por reducción del número de contagiados, sino simplemente al proceso de estimación de la función de densidad de probabilidad.

A continuación, se presenta la distribución de edades, donde se destaca que las personas entre 20 a 29 años son las que reportan mayor frecuencia de contagios seguido por las personas de 30 a 39 y de 50 a 59 años de edad.

Gráfico 5 Distribución por Edades de Personas Contagiadas



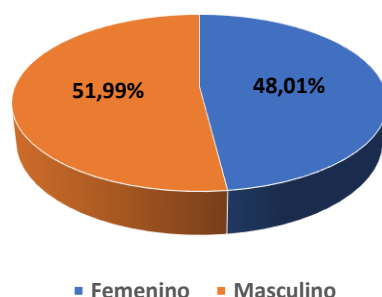
Fuente: Elaboración Propia con Información de INS (2020).

Los grupos de edad con menor nivel de contagio (a la fecha), comprenden entre los 60 a 69 (8%), 70 a 79 (5%), 10 a 19 (4%), 80 a 89 (2%), 0 a 9 años (0.3%) respectivamente, lo que preliminarmente indica es que el virus está llegando a todos los diversos grupos de edades, aunque la ventaja es que el grupo de riesgo principal comprendido por la mayoría de edad, no está dentro de los principales grupos infectados. Sin embargo, se destaca que un 8% de la población se encuentra en el

rango de los 60 a 69 y tiene reportes del virus lo que podría comprometer la vida por pertenecer al grupo de riesgo.

En relación con el sexo de los contaminados se encuentra una distribución similar como se representa a continuación.

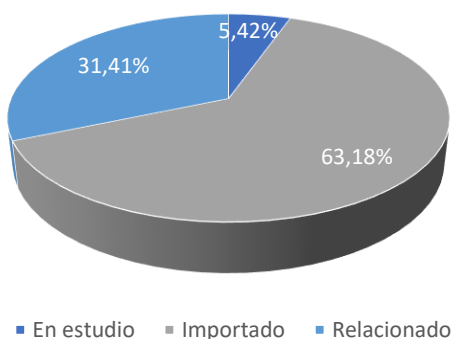
Gráfico 6 Distribución por Sexo del Total contagiados



Fuente: Elaboración Propia con Información de INS (2020).

El origen del contagio se presenta a continuación, y es de hecho la causa primaria del nivel de infección presente en el país. Se destaca que el 63.18% de los infectados con Covid-19 se debió a que lo contrajo directamente fuera del país (en especial por la predominancia por migración de Europa), mientras que el otro grueso (31.41%) fue de relacionado con las personas portadoras del virus que lo contrajeron en el exterior.

Gráfico 7 Proporción de Contagiados según origen

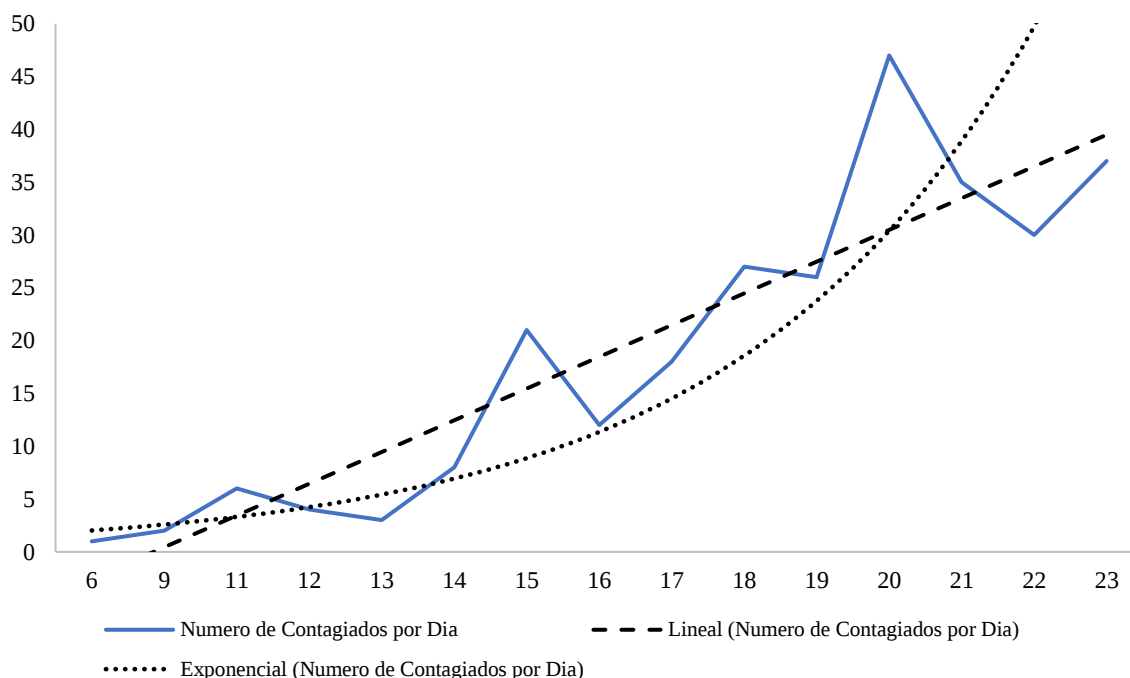


Fuente: Elaboración Propia con Información de INS (2020).

Análisis de Series Temporales

El comportamiento del número de personas contagiadas por día se presenta a continuación de acuerdo con los datos suministrados por el Instituto Nacional de Salud (2020) entre el 6 de marzo de 2020 al 23 del mismo mes y año.

Gráfico 8 Comportamiento de Contagios por día en Colombia desde el 6 de marzo de 2020



Nota: La tendencia temporal tanto de forma lineal como de su comportamiento exponencial se presenta en el gráfico. Para ver las extrapolaciones referirse al Anexo A, aunque se advierte que dichas extrapolaciones asumen una tendencia determinística, lo cual es irreal en todo el proceso de expansión. Fuente: Elaboración Propia con Información de INS (2020).

La interpretación es que existe un proceso creciente de la serie de tiempo a lo largo de los días en que se registra la información, dicho proceso si se mira con la óptica de la tendencia lineal, parece registrar un fenómeno estacional o cíclico. Es necesario demarcar que por el momento no se presenta un cambio estructural en la serie de tiempo, esto es debido a que la medida de cuarentena a nivel nacional se implementó oficialmente el 22 de marzo de 2020 a través del Decreto 457 de 2020 del Ministerio del Interior (Ministerio del Interior, 2020). Tras dicha implementación, en principio debe de existir un cambio estructural que modifique la tendencia (bien sea lineal o exponencial) del número de contagiados con el Covid-19.

Correlaciones y Estadísticas

En este apartado se presentan breves análisis de la correlación aparentemente establecida entre el número de casos confirmados por contagio de Covid-19 y las variables disponibles de acuerdo con los datos del INS (2020). Las estadísticas descriptivas se presentan a continuación:

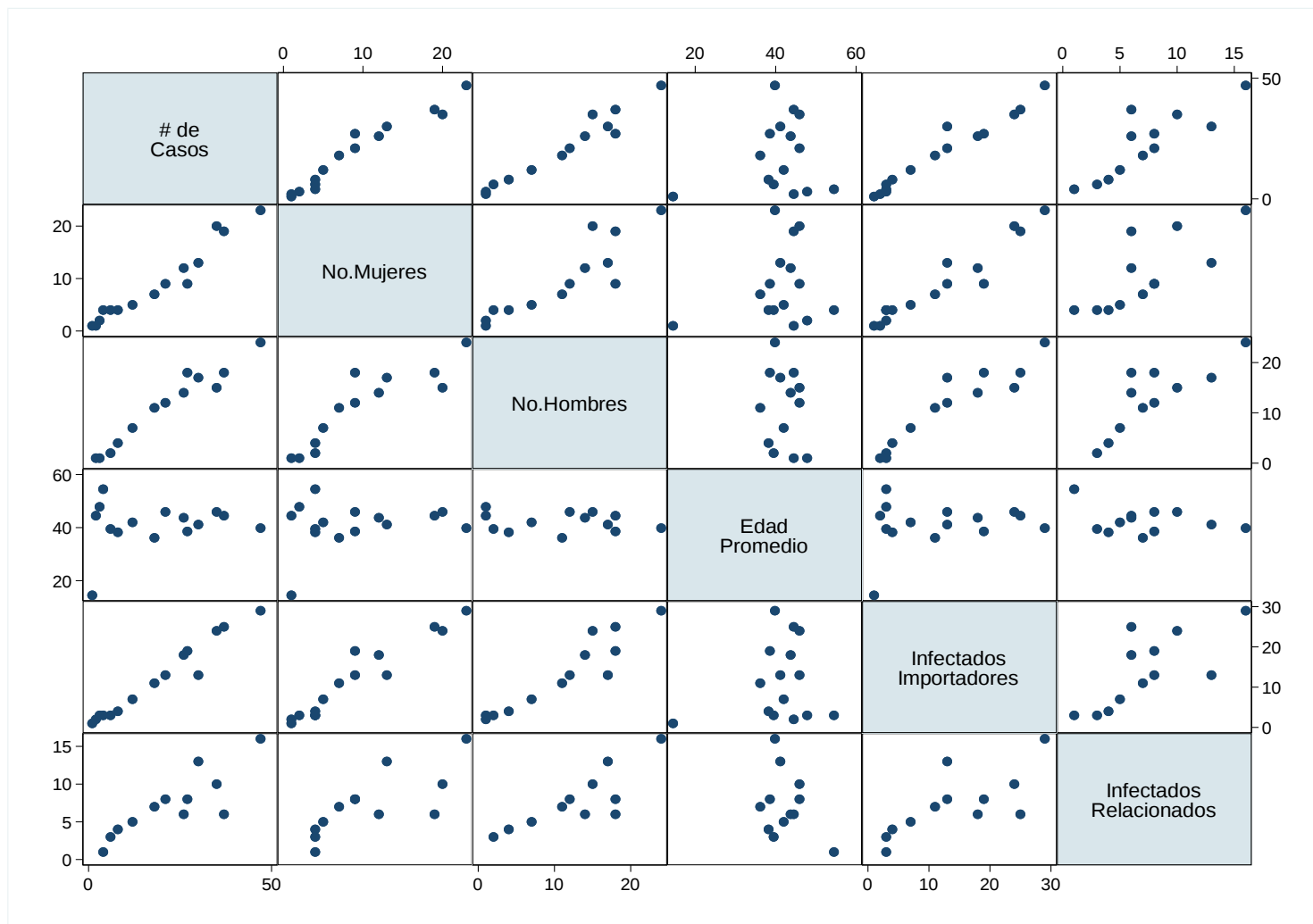
Tabla 2 Estadísticas Descriptivas de las Series Temporales

Variable	Obs	Mean	Std.Dev.	Min	Max
Número de Casos Confirmados	15	18.467	14.716	1	47
Diarios					
Número de Mujeres contagiadas por día	15	8.867	7.14	1	23
Número de Hombres contagiados por día	13	11.077	7.5	1	24
Edad Promedio de contagios por día	15	41.127	8.673	14.5	54.5
Cantidad de Infectados por Importe de virus en el extranjero.	15	11.667	9.416	1	29
Cantidad de Infectados por relación no importada	12	7.25	4.202	1	16

Nota: Los cálculos para variables para cada día se realizaron a partir de la información existente de corte transversal, en total se tienen 15 puntos temporales de acuerdo a la información obtenida. Fuente: Elaboración Propia con Información de INS (2020).

Se destaca la media de contagios diarios es de 18 personas, con un mínimo de un caso al día correspondiente al primer infectado y un máximo de 47 infectados al día correspondiente al 20 de marzo de 2020. Hay más infectados de género masculino en promedio en relación con las mujeres (11 hombres y 9 mujeres infectadas en promedio). La edad promedio a partir de los grupos de edades considerando su frecuencia de contagios es de 41 años, así mismo la cantidad promedio de infectados por haber importado al virus es superior que la de infectados que estuvieron relacionados, pero no lo importaron directamente, se espera que esta situación cambie drásticamente en el futuro.

Gráfico 9 Matriz de Correlaciones entre el Número de Casos y variables relacionadas



Fuente: Elaboración Propia con Información de INS (2020).

El patrón de correlaciones en relación con el número de casos confirmados en el tiempo, se encuentra dado positivamente tanto por el número de mujeres como el de hombres confirmados (resultado lógico puesto que entre hombres y mujeres comprenden la cantidad de personas contagiadas de forma total). Un aspecto interesante es que no parece existir una correlación ni positiva ni negativa en relación con la edad promedio y el número de contagios, así mismo se encuentra una correlación positiva para los infectados tanto relacionados como importadores del virus.

La asociación lineal positiva parece enfocarse de forma más fuerte en los importadores del virus a Colombia, en cambio, los infectados por relación indirecta tienen mayor dispersión en relación con el número de contagiados y su asociación lineal no es tan notable.

La edad promedio en relación con el resto de variables de análisis no tiene un patrón significativo, y parece ser totalmente independiente en términos de correlaciones. Sin embargo, en un estudio de mortalidad (diferente a este que estamos realizando que es de contagio) los resultados serán diferentes puesto que la edad es condición influyente para la precipitación de síntomas más agresivos.

Análisis de regresión preliminares.

Aunque el fenómeno de número de contagiados por día apenas está empezando en Colombia, es útil considerar formas de entender el fenómeno para su mitigación y control. Un inconveniente evidente es la baja cantidad de observaciones a la fecha, dado que el fenómeno no se ha logrado expandir en el tiempo, sin embargo, aquí partimos de las consideraciones de Speed (1994) en el que el bajo tamaño muestral puede ser objeto de análisis en el contexto de la regresión multivariada a través de técnicas de revalidación muestral. Estas técnicas se refieren primordialmente a los enfoques Bootstrap¹ y Jackknife en el contexto de la regresión múltiple, bajo esta idea es posible concluir inferencias acerca del número de contagiados y de su significancia estadística.

De manera ad-hoc y considerando las variables sobre las que se tiene información, se pueden plantear aproximaciones para estudiar el fenómeno de personas contagiadas en Colombia, donde se empieza definiendo tres tipos de modelos generales, el primero de regresión con las variables en niveles, el segundo con las variables en primeras

¹ Sin embargo, es de considerar que la técnica Bootstrap es sensible respecto a su significancia estadística en muestras bajas, tal y como se demuestra en los experimentos MonteCarlo con diversos tamaños muestrales (Riveros Gavilanes, 2020),

diferencias y el tercero, desde su comportamiento autorregresivo. Las ecuaciones que expresan estas ideas asumiendo una forma lineal se presentan a continuación:

$$NC_t = c + \theta t + \beta_1 MI_t + \beta_2 EP_t + \beta_3 II_t + \beta_4 IR_t + u_t \quad (1)$$

$$\Delta NC = c + \beta_1 \Delta MI + \beta_2 \Delta EP + \beta_3 \Delta II + \beta_4 \Delta IR + u_t \quad (2)$$

$$NC_t = c + \gamma_1 NC_{t-1} + u_t \quad (3)$$

Donde:

NC_t es el número de personas contagiadas en el tiempo t (con periodicidad diaria).

c es la constante o intercepto autónomo del número de contagiados, t es la tendencia lineal de la serie de personas contagiadas.

MI es el número de mujeres con coronavirus por día².

EP es la edad promedio de los infectados en un determinado día.

II es la cantidad de personas contaminadas en el extranjero (que importaron el virus al país).

IR es la cantidad de personas que se contagiaron por relación alguna manera con los que importaron el virus.

NC_{t-1} es el número de personas contagiadas en el tiempo $t - 1$ (con periodicidad diaria se refiere al día anterior).

$u_t \sim N(0, \sigma)$ es el residual de la regresión que debe satisfacer estrictamente los supuestos de normalidad, ausencia de correlación serial, homocedasticidad, no multicolinealidad perfecta, y ausencia de variables omitidas. La ecuación (1) se realiza en niveles considerando los valores contemporáneos diarios de la variable, la ecuación (2) en primeras diferencias se relaciona con la variación de las variables, esto para observar de facto el comportamiento en términos de variaciones, la ecuación (3) es el modelo con aproximación autorregresiva, donde el número de rezagos del proceso autorregresivo se estableció con los criterios de información Akaike -AIC- y Bayesiano -BIC-.

² La inclusión de esta variable permite dimensionar si el contagio se encuentra asociado estadísticamente con el sexo, sin embargo, no se puede incluir el número de hombres contagiados en el modelo puesto quedaría un caso de multicolinealidad perfecta, esto debido a que la suma de los valores de los sexos daría como resultado el total de contagiados y la regresión sería inviable.

Aquí hay serias consideraciones que necesario recalcar: 1) El número de las observaciones es insuficiente como para proceder con análisis de relaciones de largo plazo, en especial, que estamos manejando un contexto de corto plazo, 2) los modelos no están basados en condiciones, explicaciones, teoría o métodos disciplinares de la epidemiología. El proceso de obtención de información se asemeja más que nada a la técnica de minería de datos con la información disponible en relación con el número de personas contagiadas que suministra el INS, 3) Los modelos no tienen fines predictivos, su principal objetivo es observar las relaciones derivadas de la significancia estadística con las variables disponibles.

El esquema general del proceso se presenta a continuación y comprende la siguiente metodología para este caso.

Gráfico 10 Metodología del Análisis de Series Temporales



Nota: Esta es la ruta de la metodología para el análisis con las variables disponibles en relación con el número de personas contagiadas por día, al final los resultados se presentan con las tres técnicas Jackknife, Bootstrap y Mínimos Cuadrados Ordinarios. Fuente: Elaboración Propia.

Resultados

Los resultados de los modelos estimados en las ecuaciones (1), (2) y (3) se presentan a continuación en la siguiente tabla. Se destaca que los modelos estimados por mínimos cuadrados de cada especificación cumplen con los supuestos clásicos de la regresión lineal (Véase Anexo B) referidos a ausencia de variables omitidas, homocedasticidad, no autocorrelación serial, y normalidad en los residuales.

Gráfico 11 Salidas de Regresiones con Diferentes Estimaciones.

Tipo de Estimación ->	MCO			Bootstrap			Jackknife		
Ecuación ->	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
VARIABLES	NC_t	ΔNC	NC_t	NC_t	ΔNC	NC_t	NC_t	ΔNC	NC_t
MI	0.326 (0.218)			0.326 (11.18)			0.326 (0.272)		
EP_t	-0.0247 (0.0982)			-0.0247 (2.077)			-0.0247 (0.112)		
II_t	0.789*** (0.141)			0.789 (2.551)			0.789** (0.279)		
IR_t	0.774*** (0.168)			0.774 (17.58)			0.774 (0.471)		
Linear Trend (t)	0.187** (0.0716)			0.187 (4.091)			0.187 (0.193)		
ΔMI		0.0818 (0.275)			0.0818 (0.880)			0.0818 (0.367)	
ΔEP		-0.0469 (0.0901)			-0.0469 (0.0625)			-0.0469 (0.0926)	
ΔII		0.929*** (0.173)			0.929* (0.486)			0.929*** (0.210)	
ΔIR		0.958*** (0.134)			0.958*** (0.309)			0.958*** (0.289)	
NC_{t-1}			0.773*** (0.204)			0.773*** (0.176)			0.773*** (0.227)
Constant C	-0.842 (4.559)	0.725 (0.554)	7.060 (4.843)	-0.842 (58.99)	0.725 (0.970)	7.060* (3.672)	-0.842 (6.389)	0.725 (1.022)	7.060 (4.283)
Observations	12	10	12	12	10	12	12	10	12
R-squared	0.995	0.986	0.590	0.995	0.986	0.590	0.995	0.986	0.590

Nota: Valores de Significancia estadística *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. # de Replicaciones de la estimación Bootstrap 50, # de Replicaciones de la estimación Jackknife (12) para cada modelo. Fuente: Elaboración Propia

El impacto de las variables independientes es igual para todos los modelos estimados de las ecuaciones (1), (2) y (3), sin embargo, existen consideraciones muy diferentes en cuanto a su significancia estadística.

Analizando el modelo expresado en la ecuación (1), se destaca que las únicas variables que bajo un nivel del 5% de significancia estadística por mínimos cuadrados son relevantes para explicar el número de contagios son el número de personas importaron el virus desde el extranjero II_t y el número de personas contagiadas por relación IR_t . Bajo el enfoque del método Bootstrap ninguna de las variables resultó relevante (sin embargo, este método es sensible a las pruebas de significancia estadística en bajos tamaños muestrales). Finalmente, para las estimaciones con el enfoque de Jackknife, el número de personas importadores del virus II_t son estadísticamente significativos para explicar el número de contagios por día en el país.

Analizando el modelo de diferencias expresado en la ecuación (2), se destaca que las variables que poseen significancia estadística a un 5% para explicar la variación en el número de contagios son la variación en el número de personas importaron el virus desde el extranjero ΔII y el número de personas contagiadas por relación ΔIR . Estos resultados son robustos a través de las diferentes estimaciones salvo en la estimación Bootstrap en la cual ΔII es significativa al 10%.

Analizando el modelo autorregresivo de orden uno expresado en la ecuación (3) se destaca que su primer rezago es significativo al 1% para explicar el número de personas contagiadas de forma actual. La particularidad por el coeficiente de la regresión, que es menor que 1, indica que este modelo en el largo plazo tenderá a decrecer la existencia de infecciones en el tiempo.

Las variables que no tuvieron influencia en el número de contagiados por día fueron tanto la variable de sexo (que mide la cantidad de mujeres que contrajeron la infección) como la edad promedio, estos resultados son consistentes con la idea de que el contagio no discrimina sexo o grupo etario, y que la relevancia causal está determinada por el contacto directo o indirecto con portadores del Covid-19.

Conclusiones

La etapa de prevención es vital que se rija por una estricta política de cuarentena, la inferencia derivada del comportamiento tendencial sin seguir lineamientos epidemiológicos sugiere que el número de personas contagiadas al día va a tender al aumento. Se espera que con la implementación del Decreto 457 de 2020 del Ministerio del Interior (Ministerio del Interior, 2020) suceda un cambio estructural en el comportamiento de personas que se están contagiando diariamente. Se requiere un seguimiento mayor a través del tiempo sobre el brote y mortalidad del Covid-19 en Colombia.

No hubo evidencia estadística suficiente que empíricamente correlacionara el sexo o niveles de edad de los contaminados en relación con su condición de contagio del Covid-19. Sin embargo, las variables de contacto directo tanto por importar el virus del extranjero como de adquisición del

virus relacionada de forma indirecta son estadísticamente significantes para explicar el número de personas contagiadas día a día.

El impacto más significativo por el momento, lo tiene la variable de personas que importaron el virus desde el extranjero en comparación con la cantidad de personas contagiadas de forma indirecta, aunque, esto es probable que cambie a lo largo del tiempo (cuando aumente el número de personas contagiadas por relación indirecta), es necesario que las personas que estuvieron en vuelos internacionales entre diciembre de 2019 y marzo de 2020 tengan la responsabilidad y consideración con la posibilidad de ser importadores del virus. Por el momento desde el 6 de marzo de 2020, el impacto más significativo es por parte de estos importadores en la propagación de la infección de forma diaria.

Los modelos aquí presentamos no deberían considerarse para efectos de pronóstico, puesto que el enfoque principal se basa en la observación de variables cuya información está disponible a través de la recolección de datos del Instituto Nacional de Salud en Colombia.

Existe una mención especial a la situación del departamento del Valle en Colombia, el cual, a través de la estimación de densidad de probabilidad Gaussiana, demostró que puede incurrir en una mayor probabilidad de incrementar sus contagios día a día, por lo que es necesario una revisión de las políticas de contención en este departamento.

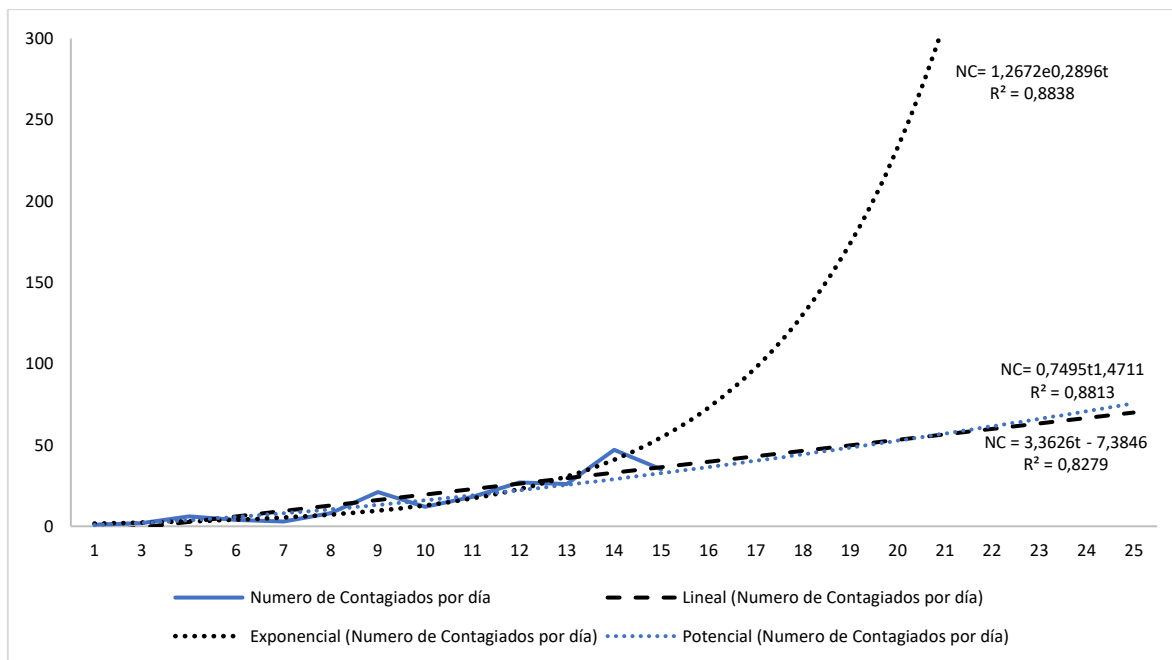
Se deben mejorar los seguimientos de los casos infectados, así mismo se recomienda crear una base de datos que pueda establecer las potenciales personas de contraer el contagio, tanto por estar en el extranjero, en Europa o Asia, desde noviembre de 2019 hasta la presente fecha, y establecer su conexión y relacionamiento con personas en términos de controlar la propagación. Esto debido a que la tasa de crecimiento del contagio se estaba incrementando abismalmente desde que se registró el primer caso el 6 de marzo. La recomendación general es que también se suspendan o se reduzca el flujo de vuelos internacionales que entran al país, dado que el impacto más significativo en estas primeras fases está dado por los importadores del virus.

Bibliografía

- Adhikari, S., Meng, S., Wu, Y., Mao, Y. P., Ye, R. X., Wang, Q. Z., . . . Zhou, H. (2020). Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. *Infectious Diseases of Poverty, Volume 9, Number 29*, 1-50.
- Bryson Taylor, D. (2020). *A Timeline of the Coronavirus*. Retrieved from The New York Times: <https://www.nytimes.com/article/coronavirus-timeline.html>
- CSSE. (2020). *Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE)*. Retrieved from John Hopkins: <https://coronavirus.jhu.edu/map.html>
- Davidson, H. (2020). *First Covid-19 case happened in November, China government records show - report* . Retrieved from The Guardian: <https://www.theguardian.com/world/2020/mar/13/first-covid-19-case-happened-in-november-china-government-records-show-report>
- Dong, Y., Mo, X., Hu, Y., Qi, X., Jiang, F., Jiang, Z., & Tong, S. (2020). Epidemiological characteristics of 2143 pediatric patient with 2019 coronavirus disease in China. *Pediatrics*, DOI: 10.1542/peds.2020-0702, 1-30.
- INS. (2020, Marzo 23). *Coronavirus en Colombia*. Retrieved from Instituto Nacional de Salud: <https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx>
- Ministerio del Interior. (2020). *Decreto Número 457 de 2020*. Retrieved from Presidencia de la República: <https://id.presidencia.gov.co/Paginas/prensa/2020/Para-garantizar-derecho-vida-salud-supervivencia-durante-Aislamiento-Preventivo-Obligatorio-Decreto-457-permite-dere-200323.aspx>
- Mizumoto, K., & Chowell, G. (2020). *Estimating Risk for Death from 2019 Novel Coronavirus Disease, China, January–February 2020*. Retrieved from Volume 26, Number 6—June 2020. Center for Disease Control and Prevention -CDC-: https://wwwnc.cdc.gov/eid/article/26/6/20-0233_article
- Riveros Gavilanes, J. M. (2020). Low sample size and regression: A Monte Carlo approach. *Journal of Applied Economic Sciences, Volume XV, Spring, 1(67)*, 22-46.
- Speed, R. (1994). Regression Type Techniques and Small Samples: A Guide to Good Practice. *Journal of Marketing Management, Vol 10, 1994*, 89-104.

Anexo A

Gráfico 12 Extrapolación de contagios con tendencia determinística (no estocástica).



Nota: De acuerdo con la fórmulas de la tendencia, para el día 25 de contagios, se calcula de acuerdo a la tendencia exponencial un total de 1767 contagios, en la tendencia lineal se calculan 77 contagiados y la tendencia potencial 85. Sin embargo, es de aclarar que estos pronósticos se dan con el supuesto irreal de tendencia determinística y sin cambio estructural. Sin embargo, la verdadera razón subyacente de no usar estos como pronósticos es que la tendencia de contagio en realidad es estocástica. Por lo que los contagios no se dan por el tiempo en si mismo, sino de los factores atenuantes a los contagios. Aquí, los estudios epidemiológicos son mucho más precisos y reales, Véase Dong et. al (2020) Adhikari et. al (2020)

Anexo B

Revisión de Supuestos del Modelo en Niveles

Tabla 3 Revisión de Especificación Correcta (No Variables Omitidas por test de Ramsey)

Ramsey RESET test using powers of the fitted values of dependent variable		
Ho: model has no omitted variables		
F(3, 3) =	0.58	
Prob > F =	0.6667	

Fuente: Elaboración Propia

Tabla 4 Revisión de Supuestos de Heterocedasticidad Modelo en Niveles Ecuación (1)

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity		
Ho: Constant variance, Variables: fitted values of dependent variable		
chi2(1) =	3.07	
Prob > chi2 =	0.0798	

Fuente: Elaboración Propia

Tabla 5 Revisión de Supuestos de Correlación Serial, Modelo en Niveles Ecuación (1)

Breusch-Godfrey LM test for autocorrelation		
H0: no serial correlation at lag (1)		
chi2(1) =	2.367	
Prob > chi2 =	0.1239	

Fuente: Elaboración Propia

Tabla 6 Revisión de Supuestos de No Multicolinealidad Perfecta

Variance Inflation Factor Test		
Variable	VIF	1/VIF
NoMujeres	14.58	0.068584
Infec_Import	10.94	0.091428
linear_trend	3.65	0.273683
Infec_Relac~o	3.37	0.296549
Edad_Promed~o	1.57	0.637149
Mean VIF	6.82	

Fuente: Elaboración Propia

Tabla 7 *Revisión de Supuestos de Normalidad Modelo en Niveles Ecuación (1)*

Skewness/Kurtosis tests for Normality
<u>H0: no serial correlation at lag (1)</u>
adj chi2(2) = 5.89
Prob > chi2 = 0.0527
Fuente: Elaboración Propia

Revisión de Supuestos del Modelo en Primeras Diferencias

Tabla 8 *Revisión de Especificación Correcta (No Variables Omitidas por test de Ramsey)*

Ramsey RESET test using powers of the fitted values of dependent variable
<u>H0: model has no omitted variables</u>
F(3, 2) = 0.56
Prob > F = 0.6919
Fuente: Elaboración Propia

Tabla 9 *Revisión de Supuestos de Heterocedasticidad Modelo en Primeras Diferencias Ecuación (2)*

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
<u>H0: Constant variance, Variables: fitted values of dependent variable</u>
chi2(1) = 2.19
Prob > chi2 = 0.1388
Fuente: Elaboración Propia

Tabla 10 *Revisión de Supuestos de Correlación Serial, Modelo en Primeras Diferencias Ecuación (2)*

Breusch-Godfrey LM test for autocorrelation
<u>H0: no serial correlation at lag (1)</u>
chi2(1) = 0.102
Prob > chi2 = 0.7489
Fuente: Elaboración Propia

Tabla 11 Revisión de Supuestos de No Multicolinealidad Perfecta

Variance Inflation Factor Test		
Variable	VIF	1/VIF
D_NoMujeres	7.35	0.136118
D_Infec_Im~t	6.54	0.153015
D_Infec_Re~o	1.62	0.619155
D_Edad_Pro~o	1.28	0.783619
Mean VIF	4.19	

Fuente: Elaboración Propia

Tabla 12 Revisión de Supuestos de Normalidad Modelo en Primeras Diferencias Ecuación (2)

Skewness/Kurtosis tests for Normality	
H0: no serial correlation at lag (1)	
adj chi2(2) =	0.47
Prob > chi2 =	0.7899

Fuente: Elaboración Propia

Revisión de Supuestos del Modelo AR (1)

Tabla 13 Selección de Rezagos Ideales por Criterios de Información

Selection-order criteria

Sample: 15mar2020 - 23mar2020

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-33.518				125.679	7.671	7.623	7.69
1	-31.923	3.190	1	0.074	110.841*	7.53844*	7.44386*	7.58226*
2	-31.349	1.148	1	0.284	124.176	7.633	7.491	7.699
3	-30.588	1.523	1	0.217	136.304	7.686	7.497	7.774
4	-30.550	0.076	1	0.783	181.953	7.900	7.664	8.010

Fuente: Elaboración Propia

Tabla 14 Revisión de Especificación Correcta (No Variables Omitidas por test de Ramsey)

Ramsey RESET test using powers of the fitted values of dependent variable	
Ho: model has no omitted variables	
F(3, 7) =	0.60
Prob > F =	0.6338

Fuente: Elaboración Propia

Tabla 15 Revisión de Supuestos de Heterocedasticidad Modelo AR(1) Ecuación (3)

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity		
Ho: Constant variance, Variables: fitted values of dependent variable		
chi2(1)	=	0.37
Prob > chi2	=	0.5428

Fuente: Elaboración Propia

Tabla 16 Revisión de Supuestos de Correlación Serial, Modelo AR(1) Ecuación (3)

Breusch-Godfrey LM test for autocorrelation		
H0: no serial correlation at lag (1)		
chi2(1)	=	1.598
Prob > chi2	=	0.2062

Fuente: Elaboración Propia

Tabla 17 Revisión de Supuestos de No Multicolinealidad Perfecta

Variance Inflation Factor Test		
Variable	VIF	1/VIF
L_casos	1.00	1.000000
Mean VIF	1.00	

Fuente: Elaboración Propia

Tabla 18 Revisión de Supuestos de Normalidad Modelo en AR (1) Ecuación (3)

Skewness/Kurtosis tests for Normality		
H0: no serial correlation at lag (1)		
adj chi2(2)	=	3.25
Prob > chi2	=	0.1967

Fuente: Elaboración Propia