

## Entrada de viajeros y densidad poblacional en la propagación de la COVID-19 en Cuba

### Arrival of travelers and population density in the spread of COVID-19 in Cuba

Larisa Zamora Matamoros<sup>1\*</sup> <https://orcid.org/0000-0003-2210-0806>

Nelsa María Sagaró del Campo<sup>2</sup> <https://orcid.org/0000-0002-1964-8830>

Luis Eugenio Valdés García<sup>2</sup> <https://orcid.org/0000-0003-1613-4305>

Ileana Benítez Jiménez<sup>1</sup> <https://orcid.org/0000-0002-4753-3416>

<sup>1</sup>Universidad de Ciencias Médicas, Santiago de Cuba, Cuba.

<sup>2</sup>Universidad de Oriente, Santiago de Cuba, Cuba.

\*Autor para la correspondencia. [larisa@uo.edu.cu](mailto:larisa@uo.edu.cu)

#### RESUMEN

**Introducción:** El comportamiento no homogéneo de la cantidad de casos confirmados con COVID-19 en diferentes regiones de Cuba aún no se ha esclarecido, lo cual resultaría de utilidad para la toma de decisiones en futuras epidemias en el país.

**Objetivo:** Determinar la influencia de la entrada de viajeros y la densidad poblacional sobre la distribución no homogénea de la cantidad de casos con COVID-19 por provincias en Cuba.

**Métodos:** Se desarrolló un estudio ecológico, exploratorio, de grupos múltiples, comparando las provincias cubanas según variables del nivel global y agregado, relacionadas con la cantidad de casos con COVID-19, confirmados durante la epidemia en Cuba. Se aplicó el análisis de regresión lineal múltiple para seleccionar el modelo que mejor describe el comportamiento de los datos y el análisis de clúster para visualizar la agrupación de las provincias.

**Resultados:** Se evidenció una correlación significativa entre la cantidad de casos con COVID-19 y la cantidad de viajeros con COVID-19, la cantidad total de viajeros que arribaron al país en marzo y los eventos de transmisión. En el modelo de regresión resultaron significativas la densidad poblacional y las cantidades de viajeros total y con COVID-19. El análisis de clúster reveló la formación de cuatro grupos de provincias.

**Conclusiones:** La cantidad de casos con COVID-19 por provincia se relaciona con la cantidad de viajeros que entraron al país, con y sin COVID-19, y la densidad poblacional. Se forman cuatro grupos de provincias por su similitud en los aspectos identificados en la regresión.

**Palabras clave:** COVID-19; regresión lineal múltiple; clúster; eventos de transmisión; viajeros; densidad poblacional.

#### ABSTRACT

**Introduction:** The non-homogeneous behavior of the number of COVID-19 confirmed cases in different regions of Cuba has not yet been clarified, which would be useful for decision-making in future epidemics in the country.

**Objective:** To determine the influence of the arrival of travelers and the population density on the non-homogeneous distribution of the number of COVID-19 cases by provinces in Cuba.

**Methods:** An ecological, exploratory, multiple group study was carried out, comparing Cuban provinces according to variables of the global and aggregate levels, related to the number of COVID-19 cases, confirmed during the epidemic in Cuba. Multiple linear regression analysis was applied to select the model that best describes the behavior of the data and cluster analysis to visualize the grouping of the provinces.

**Results:** A significant correlation was proved between the number of COVID-19 cases and the number of travelers with COVID-19, the total number of travelers who arrived in Cuba in March, and transmission events. In the regression model, the population density and the total number of travelers and those with COVID-19 were significant. The cluster analysis revealed the formation of four groups of provinces.

**Conclusions:** The number of cases with COVID-19 by province is related to the number of travelers who arrived in the country, with and with no COVID-19, and the population density. Four groups of provinces are formed by their similarity in the aspects identified at regression.

**Keywords:** COVID-19; multiple linear regression; cluster; transmission events; travellers; population density.

Recibido: 19/02/2020

Aceptado: 23/04/2020

## Introducción

El SARS-CoV-2 es un virus altamente infeccioso, causante de la COVID-19. En Cuba, hasta el 25 de mayo de 2020 se habían confirmado 1963 casos, con un saldo de 82 fallecidos.<sup>(1)</sup>

La distribución de la COVID-19 ha sido geográficamente desproporcionada. Ciertos factores climáticos y características de la población se cree que influyen en la transmisión, pero los estudios que examinan estos factores son limitados. Se invoca que la propagación epidémica es más rápida en entornos fríos, secos y de mayor altitud, así como entornos más urbanos.<sup>(2)</sup>

En Cuba, la cantidad de casos no ha tenido una distribución homogénea en las diferentes regiones del país, lo cual conduce a formular la siguiente interrogante ¿Qué aspectos estarán influyendo en la distribución no homogénea de la cantidad de casos con COVID-19 en el país?

Dado que el punto de partida de la epidemia en Cuba fueron viajeros procedentes de Italia y la hipótesis que se plantea en la literatura acerca de la influencia de la densidad poblacional sobre esta, conllevó a estudiar la posible relación de estos aspectos, y otros asociados a ellos, en la propagación del virus en Cuba.

Aunque los estudios ecológicos no pueden por sí mismos confirmar la causalidad de los efectos observados, han impulsado una gran cantidad de otros estudios epidemiológicos y de experimentos en condiciones controladas, que cada día dan más pistas sobre los mecanismos que conducen a un aumento en el riesgo de desarrollarse determinado evento.<sup>(3)</sup>

El objetivo de este trabajo fue determinar la influencia de la entrada de viajeros y la densidad poblacional sobre la distribución no homogénea de la cantidad de casos con COVID-19 por provincias en Cuba.

## Métodos

Este estudio se desarrolla con vistas a lograr una aproximación a la respuesta de la interrogante formulada, para lo cual se emplearon la regresión lineal múltiple (RLM) y el análisis de clúster (AC). La RLM es una técnica estadística que permite establecer la relación que se produce entre una variable dependiente continua  $Y$  (respuesta) y un conjunto de variables independientes o covariables  $(X_1, X_2, \dots, X_p)$ .<sup>(4,5)</sup>

El AC es una técnica estadística multivariante cuya finalidad es dividir un conjunto de objetos en grupos de forma que los perfiles de los objetos en un mismo grupo sean muy similares entre sí y los de los objetos de clúster diferentes sean distintos.<sup>(6)</sup>

Se desarrolló un estudio ecológico, exploratorio, de grupos múltiples, comparando varias regiones en un mismo momento, siendo las unidades de análisis 16 poblaciones geográficamente bien

delimitadas, que se corresponden con las 15 provincias en que se divide la República de Cuba, según la actual división político-administrativa, más el municipio especial de Isla de la Juventud. El período considerado para el estudio se extendió desde el 11 de marzo del presente año, en que se confirmó el primer caso en el país, hasta el 20 de mayo en que se realizó un primer corte exploratorio.

Se estudiaron cinco covariables de los niveles agregado y global. Del nivel agregado: número de viajeros que arribaron a cada provincia en marzo de 2020 (CantViajero), número de viajeros confirmados con COVID-19 (CasosViajCovid) y promedio de contactos que generó cada caso confirmado (PromContactos). Del nivel global: densidad poblacional (DensidadPob) y cantidad de eventos de transmisión hasta la fecha de cierre del estudio (EventosTrasm).

Como variable dependiente o de respuesta se consideró, del nivel agregado, el número acumulado de casos confirmados con COVID-19 (CasosCovid) por provincia hasta el 20 de mayo del presente año.

El número de viajeros se midió solo en el mes de marzo, en relación con la aparición del primer caso, el período de transmisibilidad de la enfermedad y la fecha en que se estableció la medida de restricción de la entrada de vuelos aéreos al país.

Para poder realizar una mejor interpretación de los resultados con respecto a la variable cantidad de viajeros, se consideró como unidad 100 viajeros y con respecto a la densidad poblacional se consideró como unidad 100 habitantes por Km<sup>2</sup>.

Los datos demográficos se obtuvieron a partir del anuario estadístico de salud, Cuba 2019,<sup>(7)</sup> los referentes a los viajeros del sistema integral de la dirección de inmigración y extranjería de Santiago de Cuba y los asociados a la enfermedad del tablero de las estadísticas de la COVID-19 en Cuba.<sup>(1)</sup>

Primeramente, se verificó la normalidad de las seis variables para determinar qué coeficiente utilizar en el análisis de correlación. Al no seguir las variables una distribución normal, se trabajó con el coeficiente de correlación de Spearman. Se aplicó la regresión lineal múltiple (RLM), empleando el método *backward step wise selection* (selección paso a paso hacia atrás), para seleccionar las covariables que mejor explican la variabilidad en la cantidad de casos con COVID-19 por provincias. Se generaron nuevos modelos que incorporaron las covariables significativas, sus interacciones y sus potencias hasta el orden dos, lo cual permitió identificar el mejor modelo de regresión. Por último, se aplicó el análisis de clúster para obtener grupos de provincias similares en cuanto a las covariables que resultaron significativas al aplicar la RLM.

Todo el procesamiento estadístico se realizó con el Minitab, versión 18.0.

Se mantuvo una conducta bioética acorde a lo dispuestos en la declaración de Helsinki de la Asociación Médica Mundial. Debido a que no se emplearon datos personales no se aplicó el consentimiento informado.

## Resultados

Durante el mes de marzo del presente año entraron a Cuba un total de 313 949 viajeros procedentes de más de 177 países, de ellos 153 se confirmaron como casos de COVID-19. A partir de estas entradas, la epidemia se propagó por todo el país, generándose, hasta el 20 de mayo, un total de 1 916 casos positivos, con un promedio de 17 contactos por caso y 42 eventos de transmisión.

Se obtuvo la matriz de correlación de Spearman y se realizó la prueba de significación para verificar, con un nivel de confiabilidad de 95 %, si el valor de la correlación obtenido es significativamente diferente de cero (tabla 1).

La cantidad de casos con COVID-19 se correlacionó de forma significativa y directa con la cantidad de viajeros con COVID-19 que entraron al país, con la cantidad de eventos de transmisión y con la cantidad de viajeros en general que entraron al país, es decir, si las tres últimas covariables aumentan su valor también lo hará la cantidad de casos confirmados con COVID-19 y viceversa.

La cantidad de casos con COVID-19 no se correlacionó, significativamente, con el promedio de contactos de los casos positivos, ni con la densidad poblacional de las provincias.

Las correlaciones entre las covariables resultaron significativas y directas solo entre los eventos de trasmisión y las cantidades de viajeros en general y con COVID-19, que entraron al país en el mes de marzo, y entre la cantidad de viajeros que entraron al país y la cantidad de ellos con COVID-19.

**Tabla 1 - Matriz de correlaciones según Spearman**

		CasosCovid	CasosViajCovid	PromContactos	EventosTrasm	DensidadPob
CasosViajCovid	Correlation Coefficient	0,655				
	Valor p	0,006*				
PromContactos	Correlation Coefficient	-0,342	-0,444			
	Valor p	0,195	0,085			
EventosTrasm	Correlation Coefficient	0,897	0,790	-0,335		
	Valor p	0,000*	0,000*	0,205		
DensidadPob	Correlation Coefficient	0,016	0,034	0,422	0,043	
	Valor p	0,953	0,901	0,104	0,876	
CantViajero	Correlation Coefficient	0,840	0,766	-0,308	0,900	0,062
	Valor p	0,000*	0,001*	0,246	0,000*	0,820

\*correlaciones significativas al 5 % ( $p < 0,05$ )

Al aplicar el método de selección paso a paso hacia atrás, se obtuvo que el mejor modelo de regresión fue:

$$CasosCovid = 0.24 * CantViajero + 3.42 * CasosViajCovid + 15.24 * DensidadPob \quad (1)$$

Con este resultado se procedió a estudiar un conjunto de modelos, en los cuales intervenían todas o algunas de las covariables del modelo.<sup>(1)</sup> En estos nuevos modelos se consideraron las interacciones entre las covariables y sus potencias, hasta el orden 2. De todos esos modelos, a continuación, se presentan los 3 mejores:

$$CasosCovid = 0.34 * CantViajero + 7.61 * CasosViajCovid \quad (2)$$

$$CasosCovid = 0.30 * CantViajero + 18.40 * DensidadPob \quad (3)$$

$$CasosCovid = 6.99 * CasosViajCovid + 19.18 * DensidadPob \quad (4)$$

### Bondad de ajuste de los modelos

Para analizar la bondad de ajuste de los modelos se emplearon un conjunto de indicadores: la desviación estándar de la distancia entre los valores de los datos y los ajustados (S), los

porcentajes de variación en la variable respuesta que es explicada por el modelo ( $R^2$ ) y ajustado para el número de covariables relativo al número de observaciones ( $R^2_{ajus}$ ), el  $R^2$  pronosticado ( $R^2_{pron}$ ), el cual permite determinar qué tan bien el modelo predice la respuesta para nuevas observaciones y un índice CpMallow para ayudar a elegir entre múltiples modelos de regresión anidados (tabla 2).

El  $R^2$ ,  $R^2_{ajus}$  y  $R^2_{pron}$  toman valores entre 0 y 100 %, mientras más grande sea su valor, el modelo se ajustará mejor a los datos. En el caso de los indicadores S y CpMallow, mientras más bajo sean sus valores, el modelo se ajustará mejor a los datos.

**Tabla 2 - Indicadores de bondad de ajuste de los modelos (1)-(4)**

Modelo	S	$R^2$	$R^2_{ajus}$	$R^2_{pron}$	CpMallow
1	25,94	99,09	98,88	98,23	1,04
2	60,33	94,70	93,94	74,04	2,00
3	33,67	98,35	98,11	80,48	2,00
4	49,51	96,43	95,92	94,83	2,00

Los cuatro modelos explicaron más de 93 % de la variabilidad contenida en la variable dependiente, con una capacidad de predicción superior al 80 % en 3 de ellos. Como el objetivo del estudio no fue encontrar un modelo para realizar predicciones, este último indicador, aunque se analizó, no fue considerado para seleccionar el mejor modelo.

### Análisis de varianza

La tabla del análisis de varianza (ANOVA) informa acerca de la existencia o no de relación significativa entre las variables. El estadístico F del ANOVA contrasta la hipótesis nula de que el valor poblacional de  $R^2$  es igual a cero y, por tanto, permite decidir si existe relación lineal significativa entre la variable dependiente y el conjunto de covariables. En la tabla 3 se presenta un resumen del ANOVA para cada modelo, los resultados completos se pueden consultar en las tablas A1 - A4 del anexo.

El valor  $p$  asociado al estadístico F, en los cuatro modelos, mostró que existe relación lineal significativa a un nivel de confiabilidad superior al 99 %. Se pudo afirmar, con este nivel de confiabilidad, que las regiones definidas por las ecuaciones (1)-(4) brindan un buen ajuste de los datos.

**Tabla 3 - Análisis de varianza**

Modelo	Fuente	GL	SC Sec.	Contribución (%)	SC Ajust.	MC Ajust.	F	p
1	Regresión	3	952 666	99,09	952 666	317 555	471,91	0,000*
	Error	13	8 748	0,91	8 748	673		
	Total	16	961 414	100				
2	Regresión	2	910 455	94,70	910 455	455 227	125,06	0,000*
	Error	14	50 959	5,30	50 959	3 640		
	Total	16	961 414	100				
3	Regresión	2	945 535	98,35	945 535	472 767	416,81	0,000*
	Error	14	15 879	1,65	15 879	1 134		
	Total	16	961 414	100				
4	Regresión	2	927 095	96,43	927 095	463 548	189,10	0,000*
	Error	14	34 319	3,57	34 319	2 451		
	Total	16	961 414	100				

\*Significativas al 1 % ( $p < 0,01$ )

## Ecuación de regresión

La tabla 4 muestra las estimaciones de los coeficientes de los modelos (Coef), el error estándar de estas estimaciones (EE coef), así como el valor de estadígrafo t de student (t) que permite docimar las hipótesis acerca de la significación o no de los coeficientes del modelo, junto con su valor de p.

Además, se muestra el valor del factor de inflación de la varianza (FIV), el cual permite detectar la existencia o no de multicolinealidad entre las covariables. Si  $FIV \geq 10$  se concluye que las covariables están correlacionadas y por lo tanto existe multicolinealidad y los resultados de las estimaciones y pruebas de hipótesis no serían válidos. En caso contrario, si  $FIV < 10$  se puede aceptar que las covariables son independientes, no existe multicolinealidad y se pueden aceptar los resultados obtenidos.

Para los datos procesados, se concluyó que existe suficiente evidencia para afirmar, con un nivel de confiabilidad superior al 99 % ( $p < 0,01$ ), que en todos los modelos sus coeficientes son diferentes de cero y, por lo tanto, las covariables asociadas a estos coeficientes ejercen una influencia significativa sobre la variable respuesta. Los valores del FIV son todos menores que 10, lo cual permitió confirmar la ausencia de multicolinealidad entre las covariables de cada modelo y con ello la validez de las estimaciones y pruebas de hipótesis realizadas.

**Tabla 4 - Coeficientes del modelo de regresión**

Modelo	Término	Coef	EE del coef.	IC de 95 %	t	p	FIV
1	CantViajero	0,24	0,04	(0,15; 0,32)	6,16	0,000*	5,14
	CasosViajCovid	3,42	1,05	(1,15; 5,69)	3,26	0,006*	6,14
	DensidadPob	15,24	1,92	(11,09; 19,40)	7,92	0,000*	4,79
2	CantViajero	0,34	0,08	(0,16; 0,51)	4,01	0,001*	4,57
	CasosViajCovid	7,61	2,11	(3,09; 12,14)	3,61	0,003*	4,57
3	CantViajero	0,30	0,04	(0,22; 0,39)	7,36	0,000*	3,57
	DensidadPob	18,40	2,16	(13,78; 23,03)	8,53	0,000*	3,57
4	CasosViajCovid	6,99	1,67	(3,41; 10,58)	4,19	0,001*	4,26
	DensidadPob	19,18	3,47	(11,75; 26,61)	5,54	0,000*	4,26

 \*Significativas al 1 % ( $p < 0,01$ )

## Análisis de los residuos

El cumplimiento de los supuestos de independencia y normalidad de los residuos se verificó por medio de las pruebas de Durbin-Watson y Anderson-Darling, respectivamente.

En la tabla A5 del anexo, se muestran los valores del estadígrafo de Durbin-Watson y los valores p asociados a la prueba de Anderson-Darling.

De la tabla de los valores críticos de la prueba de Durbin-Watson<sup>(8)</sup> se obtuvieron los valores  $d_1=0,738$  y  $d_2=1,253$  para el caso de los modelos con dos covariables, y  $d_1=0,633$  y  $d_2=1,447$  para el modelo con 3 covariables.

En los tres primeros modelos, el valor del estadígrafo de Durbin-Watson cayó en la zona de ausencia de autocorrelación ( $p=0$ ) por lo que se pudo afirmar con 99 % de confiabilidad, que en estos, los errores son independientes. En el caso particular del modelo (4) no se pudo afirmar que existe o no una autocorrelación, por caer el valor del estadígrafo en la zona de incertidumbre.

La prueba de normalidad de los residuos comprobó, con un nivel de significación de 1 % ( $p < 0,01$ ), que todos los residuales siguen una distribución Normal.

Del análisis de los residuos se concluyó que los modelos del 1 al 3 cumplen con los supuestos de normalidad e independencia.

Atendiendo al análisis integral realizado, el mejor modelo de regresión para explicar la cantidad de casos con COVID-19 en las provincias, según las covariables analizadas fue el 1:

$$\text{CasosCovid} = 0,24 * \text{CantViajero} + 3,42 * \text{CasosViajCovid} + 15,24 * \text{DensidadPob}$$

El modelo evidenció, con 99 % de confiabilidad, que la cantidad de casos con COVID-19 en el país se incrementó en 0,24 por cada 100 viajeros que arribaron a Cuba en el mes de marzo, en 3,42 por cada viajero que arribó al país con la enfermedad y en 15,24 por cada 100 habitantes por Km<sup>2</sup>.

### Análisis de clúster

Este permitió visualizar la formación de cuatro grupos de provincias (fig.). El primer grupo concentra las provincias de Pinar del Río, Santiago de Cuba, Sancti Espíritus y Camagüey, el segundo a las provincias de Artemisa, Mayabeque, Cienfuegos, Las Tunas, Guantánamo, Granma y el municipio especial Isla de la Juventud. En un tercer grupo se sitúan las provincias de Matanzas, Villa Clara, Ciego de Ávila y Holguín. El cuarto grupo solo está conformado por La Habana.

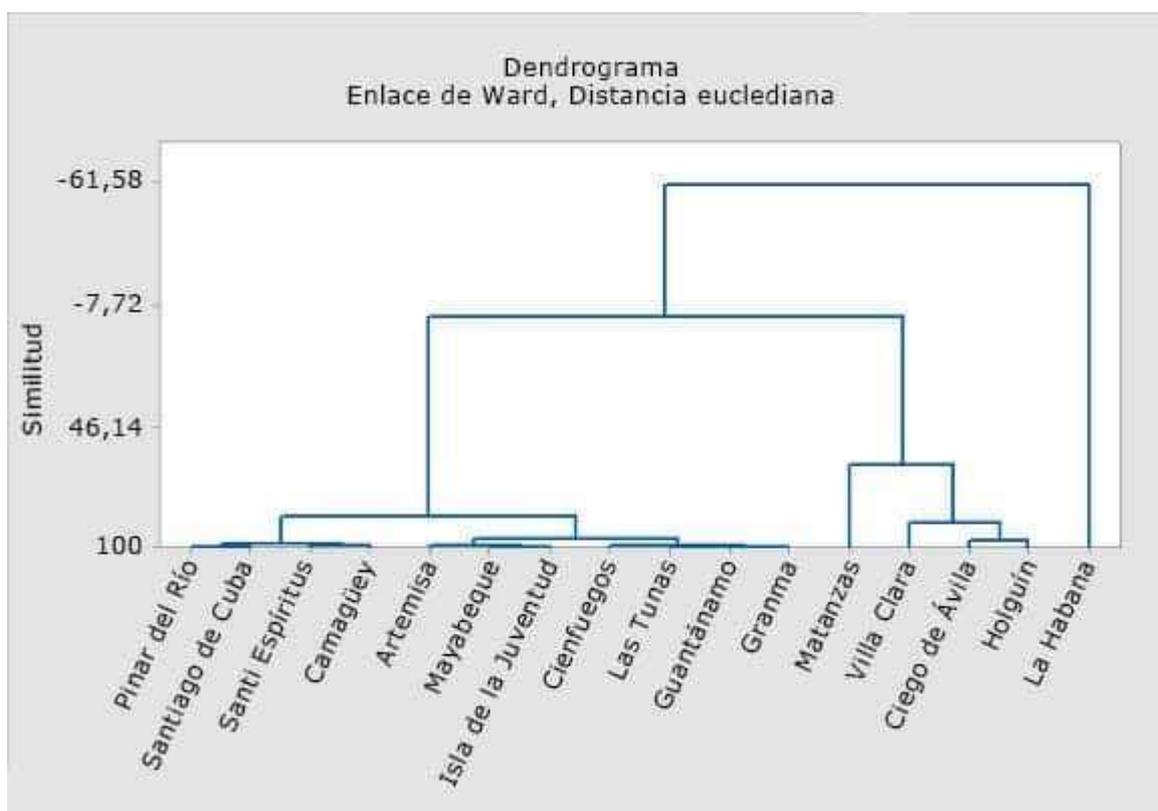


Fig. - Agrupaciones de las provincias según aspectos sociodemográficos.

Se aplicó un ANOVA teniendo en cuenta las tres covariables del modelo de RLM para verificar la hipótesis de que los cuatro grupos son significativamente diferentes. Los resultados se muestran en la tabla A6 del anexo, los cuales corroboraron la hipótesis formulada con un nivel de confiabilidad de 95 %.

## Discusión

El promedio de contactos no resultó significativo, al parecer por lo similar que fueron las acciones de salud llevadas a cabo en todo el territorio nacional, que condujeron a detectar un número promedio de contactos similar en todas las provincias.

La correlación significativa entre la cantidad de eventos de transmisión y el número de casos es lógica, ya que al existir eventos de transmisión en una comunidad aumenta el número de casos. Por ello, entre las principales estrategias de control deben estar las dirigidas a la prevención de la

COVID-19 en lugares donde se concentran grupos vulnerables como hogares maternos o de ancianos, prisiones, entre otros lugares, donde existen conglomerados de personas, como los centros de trabajo, en los cuales hay un mayor nivel de exposición. Otras medidas útiles en la prevención de la COVID-19 son la protección mediante el uso del nasobuco y la higiene con el lavado frecuente de las manos.

La eficacia de este control depende de los patrones presintomáticos y asintomáticos de la enfermedad. Con 100 % de los casos sintomáticos y un número básico de reproducción ( $R_0$ ) de 1.5 podría ser controlada la transmisión aislando 50 % de los contactos. Con un  $R_0$  de 3.5, se requiere 90 %; sin embargo, con la transmisión presintomática y potencialmente asintomática, esta efectividad se reduce aún más. En Cuba, la localización oportuna de los contactos con su correspondiente aislamiento ha sido fundamental para reducir la expansión de la epidemia. El control de los eventos de transmisión, se presupone, depende de la densidad de la población y las medidas de control, como el distanciamiento social y la cuarentena.<sup>(9)</sup> Sin embargo, esta variable no fue significativa al incluirla en el modelo de regresión, lo cual puede ser explicado por las rápidas acciones desarrolladas en Cuba, cada vez que ha aparecido un evento de transmisión, lo cual ha permitido contener el brote. Un resultado inverso ocurrió entre la densidad poblacional y el número de casos, las cuales no exhibieron una correlación significativa en el análisis bivariado, pero en el multivariado la densidad resultó ser significativa.

Dado que el virus de la COVID-19 se propaga cuando las personas se encuentran muy próximas, la densidad poblacional parece ser uno de los aspectos implicados en la propagación. Estudios como el de *Freeman* y otros<sup>(10)</sup> realizados en la ciudad de Nueva York, evalúan la influencia de la densidad poblacional en ciertas patologías. Estos autores encontraron una correlación muy baja al estudiar la densidad de manera independiente, y aún más baja cuando se tienen en cuenta otras variables de carácter socioeconómico, como el origen étnico y los ingresos económicos.

La hipótesis de que las regiones con alta densidad poblacional son más vulnerables a las epidemias, debido a la mayor posibilidad de contactos interpersonales, llevó al Grupo del Banco Mundial a averiguar si la densidad de población era un factor clave en la propagación del nuevo coronavirus, por lo que recogieron datos de 284 ciudades chinas, encontrando que las ciudades más densamente pobladas como Shanghai, Beijing, Shenzhen, Tianjin y Zhuhai, tuvieron muchos menos casos confirmados por cada 10 000 personas. Coincidentemente, estas ciudades son también las de mayor producto interno bruto, lo cual las hace más capaces de movilizar recursos para enfrentar la enfermedad y pudiera explicar, en parte, sus bajas tasas de infección. Por el contrario, las ciudades con las tasas más altas de infección fueron aquellas con densidades relativamente bajas, aunque otros factores como la cercanía a la provincia de Hubei o la gran cantidad de personas que se trasladan a Wuhan por laborar allí podrían explicar estas altas tasas, ya que se observó que la tasa de infección disminuyó a medida que aumentó la distancia desde las ciudades hasta Wuhan.<sup>(11)</sup>

Otro factor que se consideró en este estudio del Banco Mundial fue que, dado un cierto umbral de densidad de población, las ciudades brindan instalaciones y servicios de mayor calidad a sus residentes siendo más fácil para estos permanecer en casa y evitar el contacto innecesario con los demás. Así las consideraciones de este estudio están a favor del efecto beneficioso que ejerce la densidad poblacional en la lucha contra el coronavirus al plantear que la urbanización acerca a las personas tanto espacial como económicamente y que la conjunción de mayor densidad, menor distancia y mejor integración del mercado permiten un desarrollo económico que garantiza instituciones mejor diseñadas, infraestructuras de alta calidad e intervenciones eficaces para llevar a cabo el distanciamiento social que hace a las ciudades más fuertes contra las enfermedades infecciosas.<sup>(11)</sup>

Por su parte el estudio de *Verbrugge* y otros<sup>(12)</sup> consideró la relación de la densidad poblacional con actitudes sociales negativas y comportamientos indeseables. Ellos realizaron una encuesta en Baltimore para probar las hipótesis sobre los efectos positivos y negativos de la densidad poblacional y determinaron que el gran tamaño de la población y la sensación de que una zona está superpoblada producen ciertas frustraciones.

El criterio de los autores de la presente investigación es que la densidad poblacional puede influir de manera positiva o negativa, en dependencia del desarrollo socioeconómico alcanzado por las ciudades.

Con respecto a la movilidad espacial se plantea que el aumento de viajeros internacionales ha incidido en una mayor propagación de enfermedades infecciosas, nuevas y reemergentes.<sup>(13)</sup> Esto ha hecho más evidente la necesidad de prevenir y diagnosticar, oportunamente, enfermedades que no son endémicas, para lo cual es fundamental conocer los riesgos en cada destino y las poblaciones más vulnerables.<sup>(14)</sup>

Generalmente, las ciudades de mayor densidad se corresponden con un mayor desarrollo, entre ellos el de la esfera turística. En Cuba, a estas zonas de atracción entró una mayor cantidad de personas procedentes de países con altas tasas de la enfermedad, lo cual incrementó la probabilidad de que estos estuviesen infectados y propagasen el virus.

El monitoreo de los viajeros para la detección y la cuarentena de enfermos puede ser una herramienta eficaz para prevenir y retrasar la importación de algunas enfermedades. Para un control efectivo de los viajeros se precisa, además, información que ayude a los funcionarios de salud a localizar, rápidamente, a todos los viajeros en riesgo, con la realización de pruebas oportunas para detectar casos e investigar sus contactos.<sup>(15)</sup> En Cuba, todas estas medidas se implementaron tempranamente por las autoridades competentes.

El estudio de *Chinazzi*<sup>(16)</sup> midió el impacto de la limitación de la movilidad espacial en China, donde se estimó que la cuarentena de Wuhan retrasó la progresión de la epidemia de 3 a 5 días en China continental, pero tuvo un efecto más marcado a escala internacional, donde las importaciones de casos se redujeron en casi 80 % hasta mediados de febrero.

*Jin* y otros<sup>(17)</sup> atribuyeron la mayor cantidad de casos en ciertas provincias chinas a una mayor proporción de viajes a estas provincias desde Wuhan, donde se inició el primer brote de COVID-19. Otro estudio que empleó un análisis estadístico similar al de este trabajo, examinó los datos de los casos y fallecidos de COVID-19 reportados hasta el 2 de mayo de 2020 en 205 países y territorios, encontrando asociación muy significativa con el mayor porcentaje de población urbana y el aumento de viajeros procedentes del exterior, además estudiaron aspectos climáticos y geográficos, encontrando asociación con la baja temperatura, la menor humedad y la mayor altitud.<sup>(2)</sup>

Estos factores, no tenidos en cuenta en el presente estudio, pudieron también haber influido en la diseminación desigual de la enfermedad, pues las regiones más afectadas se localizan en la zona que se corresponde con la costa norte del país.

No obstante, *Pedrosa*<sup>(18)</sup> que también estudió las variables climáticas observó que estas solo eran significativas en los estudios bivariados, pero al incluirlas en modelos multivariados solo la densidad de población y la variable de la línea de tiempo mostraron significación estadística.

Se concluye que la cantidad de casos con COVID-19 por provincia se relaciona con la cantidad de viajeros que entraron al país, con y sin el virus y la densidad poblacional. Se forman cuatro grupos de provincias por su similitud en los aspectos identificados en la regresión.

## Referencias bibliográficas

1. Dashboard Cuba: Covid-19. 2020[acceso: 19/05/2020]. Disponible en: <https://covid19cubadata.github.io/#cuba>
2. Leung NY, Bulterys MA., Bulterys PL. Predictors of COVID-19 incidence, mortality, and epidemic growth rate at the country level. MedRxiv. 2020. (preprint) doi: <https://doi.org/10.1101/2020.05.15.20101097>
3. Blanco Becerra LC, Pinzón Flórez CE, Idrovo AJ. Estudios ecológicos en salud ambiental: más allá de la epidemiología Biomédica. 2015;35:191-206. Doi: <http://dx.doi.org/10.7705/biomedica.v35i0.2819>
4. Wolfgang K, Léopold S. Applied Multivariate Statistical Analysis. Fourth Edition. Springer-Verlag

- Berlin Heidelberg; 2015[acceso: 19/05/2020]. Disponible en: <https://link.springer.com/content/pdf/10.1007%2F978-3-662-45171-7.pdf>
5. Stockemer D. Multivariate Regression Analysis. In Quantitative Methods for the Social Sciences. Springer, Cham; 2019[acceso: 19/05/2020]:163-74. Disponible en: [https://link.springer.com/chapter/10.1007/978-3-319-99118-4\\_9](https://link.springer.com/chapter/10.1007/978-3-319-99118-4_9)
6. Salvador Figueras M. Análisis de conglomerados o clúster. Universidad de Zaragoza, España; 2001[acceso: 29/05/2020]. Disponible en: <http://www.5campus.org/leccion/cluster>
7. Minsap Dirección de Registros Médicos y Estadísticas de Salud Anuario de Salud 2019. La Habana; 2020[acceso: 19/05/2020]:97. Disponible en <https://salud.msp.gob.cu/?tag=anuario-estadistico-de-salud>
8. Zaiantz C. Real statistics resource pack. 2020[acceso: 19/05/2020]. Disponible en: <http://www.real-statistics.com/statistics-tables/durbin-watson-table/>
9. Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. Journal of Travel Med. 2020;27(3):038. Doi: <https://doi.org/10.1093/jtm/taaa038>
10. Freedman JL, Heshka S, Levy A. Population density and pathology: Is there a relationship? Journal of Experimental Social Psychology. 1975;11(6):539-52. Doi: [https://doi.org/10.1016/0022-1031\(75\)90005-0](https://doi.org/10.1016/0022-1031(75)90005-0)
11. World Bank Blog, Fang W, Wahba S. Urban Density Is Not an Enemy in the Coronavirus Fight: Evidence from China. 2020[acceso: 19/05/2020]. Disponible en: <https://blogs.worldbank.org/sustainablecities/urban-density-not-enemy-coronavirus-fight-evidence-china>
12. Verbrugge LM, Taylor RB. Consequences of Population Density and Size. Urban Affairs Quarterly. 1980[acceso: 19/05/2020];16(2):135-160. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/12279512>
13. Menchero Sánchez M. Flujos turísticos, geopolítica y COVID-19: cuando los turistas internacionales son vectores de transmisión. Geopolítica(s). 2020;11(Especial):105-14. Doi: <https://dx.doi.org/10.5209/geop.69249>
14. Perret PC. Infecciones en viajeros internacionales. Rev. Med. Clin. Condes. 2014;25(3):565-68. Doi: [https://doi.org/10.1016/S0716-8640\(14\)70071-9](https://doi.org/10.1016/S0716-8640(14)70071-9)
15. Myers JF, Snyder RE, Porse CC. Identification and Monitoring of International Travelers During the Initial Phase of an Outbreak of COVID-19. Morb Mortal Wkly Rep 2020;69(19):599-602. Doi: <http://dx.doi.org/10.15585/mmwr.mm6919e4>
16. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020[acceso: 19/05/2020];368(6489):395-400. Disponible en: <https://science.sciencemag.org/content/368/6489/395/>
17. Jin L, Zhao Y, Zhou J, Tao M, Yang Y, Wang X, *et al.* Distribución temporal, geográfica y por población de la nueva enfermedad por coronavirus (COVID-19) desde el 20 de enero hasta el 10 de febrero de 2020 en China. Rev Clin Esp. 2020. Doi: <https://doi.org/10.1016/j.rce.2020.04.001>
18. Pedrosa RH. The dynamics of Covid-19: weather, demographics and infection timeline. MedRxiv. 2020 (preprint). Doi: <https://connect.medrxiv.org/relate/content/181>

#### Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.

#### Contribuciones de los autores

*Larisa Zamora Matamoros*: Redacción del artículo y revisión final para la publicación.

*Nelsa María Sagaró del Campo, Luis Eugenio Valdés García y Ileana Benítez Jiménez*: Redacción del artículo.

## Anexo

Tabla A1 - Análisis de Varianza para el modelo 1

Fuente	GL	SC Sec	Contribución (%)	SC Ajust	MC Ajust	F	p
Regresión	3	952666	99,09	952666	317555	471,91	0,000
CantViajero	1	862991	89,76	25571	25571	38,00	0,000
CasosViajCovid	1	47464	4,94	7132	7132	10,60	0,006
DensidadPob	1	42212	4,39	42212	42212	62,73	0,000
Error	13	8748	0,91	8748	673		
Total	16	961414	100				

Tabla A2 - Análisis de Varianza para el modelo 2

Fuente	GL	SC Sec	Contribución (%)	SC Ajust	MC Ajust	F	p
Regresión	2	910455	94,70	910455	455227	125,06	0,000
CantViajero	1	862991	89,76	58474	58474	16,06	0,001
CasosViajCovid	1	47464	4,94	47464	47464	13,04	0,003
Error	14	50959	5,30	50959	3640		
Falta de ajuste	13	50941	5,30	50941	3919	217,70	0,053
Error puro	1	18	0,00	18	18		
Total	16	961414	100				

Tabla A3 - Análisis de Varianza para el modelo 3

Fuente	GL	SC Sec	Contribución (%)	SC Ajust	MC Ajust	F	p
Regresión	2	945535	98,35	945535	472767	416,81	0,000
CantViajero	1	862991	89,76	61413	61413	54,14	0,000
DensidadPob	1	82544	8,59	82544	82544	72,77	0,000
Error	14	15879	1,65	15879	1134		
Total	16	961414	100				

Tabla A4 - Análisis de Varianza para el modelo 4

Fuente	GL	SC Sec	Contribución (%)	SC Ajust	MC Ajust	F	p
Regresión	2	927095	96,43	927095	463548	189,10	0,000
CasosViajCovid	1	851981	88,62	42974	42974	17,53	0,001
DensidadPob	1	75115	7,81	75115	75115	30,64	0,000
Error	14	34319	3,57	34319	2451		
Total	16	961414	100				

Tabla A5 - Estadígrafo de Durbin-Watson y valores p asociados a la prueba de Anderson-Darling

Modelo	Durbin-Watson	Anderson Darlin
(1)	1,86	0,017
(2)	1,26	0,141
(3)	2,56	0,073
(4)	1,14	0,137

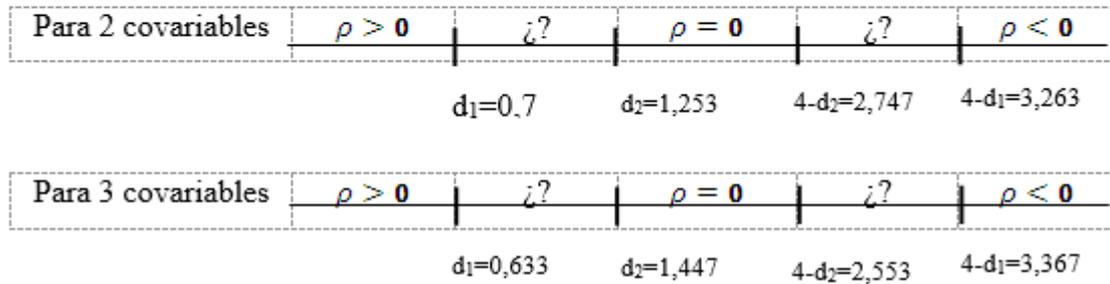


Tabla A6 - Estadígrafo F y valor de p del ANOVA

Covariable	F	p
CasosViajCovid	51,84	0,000
CantViajero	48,40	0,000
DensidadPob	1957,31	0,000