

# Un modelo multinivel para la predicción del total de casos confirmados de COVID-19 en Cuba: comparación con otros países

## A multilevel model for the prediction of the total confirmed cases of COVID-19 in Cuba: comparison with other countries

Minerva Montero Díaz <sup>1\*</sup>

**Resumen** Son muchos los indicadores que proporcionan una perspectiva global sobre la evolución de la pandemia de COVID-19 en el mundo. En este artículo, el comportamiento de la tasa del total de casos confirmados en Cuba se compara con el de otros países, utilizando modelos multinivel para datos longitudinales. El objetivo fundamental es obtener predicciones de la evolución esperada de este indicador en nuestro país. El estudio incluye un grupo de países con características demográficas similares a las de Cuba. La estrategia propuesta permite identificar un subgrupo de naciones, cuya curva media estimada del total de casos confirmados por millón de habitantes, puede servir como referente del comportamiento esperado en el país. Los correspondientes intervalos de confianza se interpretan como canales de seguimiento, que permiten identificar con claridad las posibles desviaciones en la trayectoria de la tasa de casos confirmados en Cuba, además de servir como instrumento de predicción. Suponiendo un escenario en el cual se mantenga la efectividad de las medidas de control de la COVID-19 en Cuba, se podría esperar que la tasa de casos confirmados en Cuba no exceda significativamente los límites superiores de los intervalos de confianza de las curvas medias estimadas, correspondientes a los grupos de países seleccionados para la comparación.

**Abstract** Many indicators provide a global perspective on the evolution of the COVID-19 pandemic in the world. In this paper, the behavior of the rate of total confirmed cases in Cuba is compared with that of other countries, using multilevel models for longitudinal data. The fundamental objective is to obtain predictions of the expected evolution of this indicator in our country. The proposed strategy identifies a subgroup of nations, whose estimated mean curve of the total confirmed cases per million inhabitants, can serve as a benchmark for the expected behavior in the country. The corresponding confidence intervals are interpreted as monitoring channels, which make it possible to clearly identify possible deviations in the trajectory of the rate of confirmed cases in Cuba, in addition to serving as a prediction instrument. Assuming a scenario in which the effectiveness of the COVID-19 control measures in Cuba is maintained, it could be expected that the rate of confirmed cases in Cuba does not significantly exceed the upper limits of the confidence intervals of the estimated mean curves corresponding to the groups of countries selected for comparison.

### Palabras Clave

COVID-19, Coronavirus, Modelos multinivel, Predicción

### Keywords

COVID-19, Coronavirus, Multilevel models, Prediction

<sup>1</sup> Dpto. de Matemática, Instituto de Cibernética, Matemática y Física, La Habana, Cuba, [minerva@icimaf.cu](mailto:minerva@icimaf.cu)

\*Autor para Correspondencia, Corresponding Author

## Introducción

El 11 de marzo de 2020, la Organización Mundial de la Salud (OMS) declaró a la COVID-19 como una pandemia [13]. En medio de una enorme incertidumbre sobre la evolución de la epidemia de esta enfermedad en el mundo,

diferentes modelos matemáticos, como los ya muy populares modelos SIR (Susceptibles, Infectados, Recuperados) [7], y otras extensiones de este clásico modelo [2], aportaron útiles pronósticos sobre futuros escenarios de la epidemia. Otros modelos, contruidos desde un enfoque estadístico, se han usa-

do para pronósticos a más corto plazo, utilizando aprendizaje automático o regresión [6].

En este artículo se presenta una estrategia de modelación estadística para el pronóstico a corto y mediano plazo de uno de los indicadores epidemiológicos utilizados para evaluar la evolución de la pandemia: el total de casos confirmados de COVID-19. El procedimiento se basa en el empleo de los modelos multinivel [4, 11, 8] para describir el comportamiento de este indicador en Cuba y predecir su evolución utilizando datos de otros países. Este tipo de modelo no explica la dinámica epidemiológica (como cuándo ocurrirá el pico y si ocurrirá un rebrote de la enfermedad), pero produce pronósticos cuantitativos que los encargados de formular políticas pueden necesitar a corto plazo para asignar recursos o planificar intervenciones [5]. Además, en el caso particular de los modelos multinivel, el estudio comparativo con otros países puede aprovecharse para deducir la eficacia de la intervención.

El problema se trata como un estudio longitudinal o de medidas repetidas en el tiempo [12, 10], donde los países y las ocasiones en que se mide el indicador epidemiológico se perciben como un sistema jerárquico de dos niveles. Los modelos jerárquicos de dos niveles tienen en cuenta cuánto de la heterogeneidad de los datos se debe a la evolución de la enfermedad en el tiempo y cuánto a posibles diferencias contextuales entre los países. De aquí que un examen de la heterogeneidad del total de casos confirmados de COVID-19 entre los países con presencia de coronavirus, permite hacer una valoración comparativa del ritmo de propagación de la enfermedad en Cuba. A la vez, algunos países pueden reconocerse como referentes del comportamiento esperado en nuestro país.

El objetivo de este trabajo es mostrar, cómo los modelos multinivel para datos longitudinales, capaces de examinar formas complejas de variación entre países, pueden utilizarse para la construcción de un sencillo instrumento de predicción de indicadores epidemiológicos del nuevo coronavirus.

El estudio parte de una muestra de países con características demográficas similares a las de Cuba. Para evitar el sesgo debido a las diferencias en el tamaño de la población entre los diferentes países, el análisis de los datos se basó en el número de casos confirmados por millón de habitantes [9].

Las curvas de crecimiento estimadas para cada país se utilizan como patrones de comparación con la trayectoria del indicador epidemiológico en Cuba. Se espera que la curva estimada (y la trayectoria en el tiempo de los valores observados) para Cuba no difieran significativamente de los valores medios predichos por los parámetros fijos de los modelos propuestos. En el caso de que los datos observados en Cuba se encuentre alejados de la curva media estimada, se propone adicionar al modelo una variable contextual que explique las diferencias entre los países, para estimar entonces las curvas de crecimiento específicas para los grupos de países con comportamientos homogéneos.

La curva media estimada más próxima a la evolución del indicador epidemiológico en Cuba, puede servir como

patrón de referencia del comportamiento esperado en el país. Los correspondientes intervalos de confianza (indicador del margen de error que se comete en la estimación) se interpretan como canales de seguimiento que permiten identificar con claridad las posibles desviaciones en la trayectoria de la tasa de casos confirmados en Cuba a través del tiempo, además de servir como un instrumento de predicción de la evolución esperada en el país.

En las próximas sesiones se describe el problema y los datos utilizados, se presentan algunos conceptos básicos de los modelos multinivel utilizados, y se discute, sobre la base de los resultados prácticos, las potencialidades de estos modelos en la predicción de indicadores epidemiológicos relacionados con la COVID-19.

## 1. Materiales y Métodos

### 1.1 Los datos

Para comprender la propagación de la enfermedad en Cuba es útil evaluar, entre otros muchos indicadores, con qué rapidez aumenta la tasa del total de casos confirmados de COVID-19 en comparación con el crecimiento promedio de otros países. Debido a las diferencias contextuales entre las naciones, en este estudio se determinó limitar la comparación a un grupo de países (ver Tabla 1) con características demográficas semejantes al nuestro.

**Tabla 1.** Grupo de países con características demográficas similares a Cuba seleccionados mediante técnicas de inteligencia artificial [1]. Los indicadores demográficos utilizados para la clasificación son: Población, Densidad poblacional, Tasa de Fertilidad, Edad Media y Población Urbana.

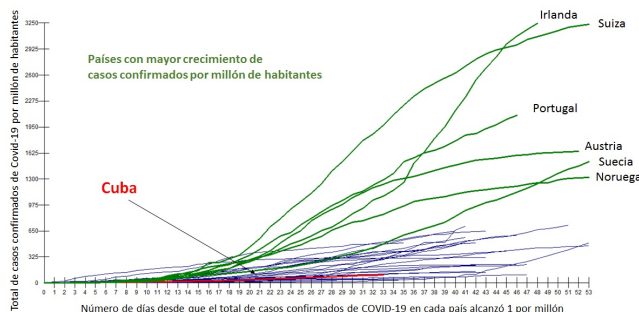
|                        |                     |
|------------------------|---------------------|
| Armenia                | Honduras            |
| Austria                | Hungría             |
| Azerbaiyán             | Irlanda             |
| Bielorrusia            | Jordán              |
| Bosnia y Herzegovina   | Letonia             |
| Brunei                 | Lituania            |
| Chile                  | Macedonia del Norte |
| Chipre                 | Moldavia            |
| Costa Rica             | Montenegro          |
| Croacia                | Noruega             |
| Emiratos Árabes Unidos | Nueva Zelanda       |
| Eslovaquia             | Portugal            |
| Finlandia              | Suecia              |
| Georgia                | Suiza               |
| Grecia                 | Túnez               |

La muestra de estudio estuvo constituida por 30 países, más Cuba. Para cada país se contó con el registro diario del total de casos confirmados por millón de habitantes. La fuente de los datos corresponde al Centro Europeo para la Prevención y Control de Enfermedades (European CDC) [3].

En este estudio, el registro del indicador epidemiológico comienza el día en que el país tenía una tasa de 1 caso por millón de habitantes. Esto permite hacer comparaciones de la rapidez con la que ha aumentado el número de casos confirmados en diferentes países [9]. En el caso de Cuba la serie de datos comienza el 20 de marzo, sexto día la epidemia en nuestro país.

El resultado es un conjunto de datos longitudinales con un número desigual de mediciones por países y observaciones perdidas, como cuando ocurre si algún país no reporta las estadísticas de algún día. No obstante, el enfoque multinivel es capaz de acomodar este tipo de datos desbalanceados y es eficiente aun cuando algunos datos se pierden aleatoriamente y los intervalos de tiempo entre mediciones pueden variar entre los países.

El análisis de los datos se realizó en dos momentos diferentes. Las fechas de actualización de los datos utilizados fueron 22 de abril y 17 de mayo de 2020. En el primer estudio, el período de análisis para Cuba comprendió 33 días. Las series de otros países comprendieron hasta 53 días. En el segundo estudio (actualización de los datos) el período para Cuba contó con 58 días. El resto de los países, hasta 72 días. El gráfico de las trayectorias hasta el día 22 de abril de la tasa del total de casos confirmados (*TTCC*) de cada uno de los 30 países, más Cuba, se muestran en la figura 1.

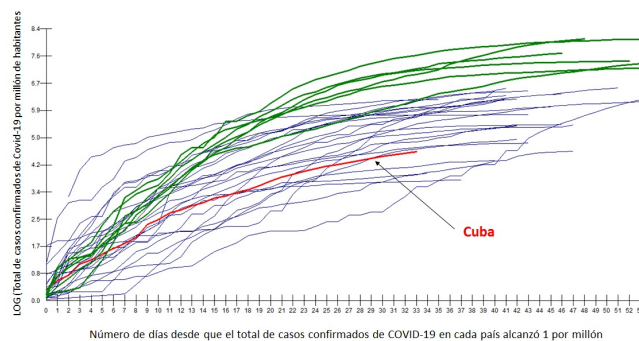


**Figura 1.** Trayectoria por países del número total diario de casos confirmados de COVID-19 por millón de habitantes

Para aproximar los datos a una distribución normal se aplicó una transformación logarítmica a la variable *TTCC*. En la figura 2 se muestra el gráfico de las trayectorias después de transformada la variable.

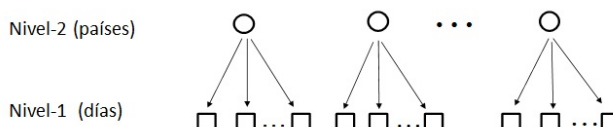
## 1.2 Modelos multinivel para datos longitudinales

Mediante modelación multinivel se obtiene un resumen del comportamiento de la variable de interés (total de casos confirmados de COVID-19 por millón de habitantes) en función del tiempo, tomando en consideración la estructura de correlación de las medidas repetidas y comparando este indicador entre la muestra de países. El primer paso es organizar los datos en un formato adecuado para el análisis. Así, se establece que la variable respuesta, registrada diariamente en cada uno de los países, da lugar a una estructura jerárquica en la que las medidas repetidas se anidan dentro de los países



**Figura 2.** Trayectoria por países del Log (Total de casos confirmados por millón de habitantes)

seleccionados de la población de países con COVID-19, de manera que las ocasiones (días) constituyen las unidades de nivel-1 y los países, las unidades de nivel-2 (Figura 3). La variación de la variable respuesta intra-países a través del tiempo ocurre en el nivel más bajo (nivel-1) y la variación de la variable respuesta media entre-los países ocurre en el nivel-2.



**Figura 3.** Estructura jerárquica de los datos.

Una ventaja importante de la modelación multinivel es que permite particionar la varianza de la variable respuesta en dos componentes: la varianza intra-países y la varianza entre-países.

### 1.2.1 Modelo polinomial

Sea  $y_{it}$  el valor de la variable respuesta del país  $i$  medida en la ocasión  $t$ ;  $(dia)_{it}$  es la variable indicadora del momento en que se hace la medición ( $i = 1, \dots, n$ ;  $t = 1, \dots, m_i$ ). Se supone que para cada uno de los países el comportamiento de la respuesta observada a través del tiempo puede describirse mediante el modelo de nivel-1:

$$y_{it} = \beta_{0i} + \beta_{1i}(dia)_{it} + \beta_{2i}(dia)_{it}^2 + \beta_{3i}(dia)_{it}^3 + e_{it}, \quad (1)$$

donde  $\beta_{0i}$ ,  $\beta_{1i}$ ,  $\beta_{2i}$  y  $\beta_{3i}$  son los coeficientes de regresión de la  $i$ -ésima ecuación de nivel-1 y  $e_{it}$  captura la variación de la respuesta en el momento  $t$  sobre la curva de crecimiento del  $i$ -ésimo país, tal que  $E[y_{it}] = \beta_{0i} + \beta_{1i}(dia)_{it} + \beta_{2i}(dia)_{it}^2 + \beta_{3i}(dia)_{it}^3$ , suponiendo que  $E[e_{it}] = 0$  y  $var[e_{it}] = \sigma_e^2$ .

Para modelar la variación entre los países se formulan ecuaciones adicionales de nivel-2 donde uno o ambos de los parámetros de regresión de nivel-1 se modelan como la suma de una media general más una desviación aleatoria de la media. Se sabe que la variable respuesta en cada ocasión varía entre los países. Esta variación queda expresada por la inclusión

del término aleatorio  $u_{1i}$ , que representa la desviación de la variable respuesta a través del tiempo en el país  $i$ , con respecto a los valores promedios de la población. El intercepto, que representa la variable respuesta en el momento inicial de la serie (cuando el total de casos alcanzó 1 por millón de habitantes en cada país) se considera un parámetro fijo. Así, para los datos del ejemplo, las ecuaciones de regresión en el nivel-2 son:

$$\beta_{0i} = \gamma_{00} ; \beta_{1i} = \gamma_{10} + u_{1i} ; \beta_{2i} = \gamma_{20} ; \beta_{3i} = \gamma_{30} \quad (2)$$

donde  $u_{1i}$  son errores aleatorios con esperanza y varianza:

$$E(u_{1i}) = 0 ; \text{var}(u_{1i}) = \sigma_{u_1}^2,$$

Este modelo de dos niveles puede escribirse como una única ecuación sustituyendo las ecuaciones (2) en la ecuación (1). Reordenando términos se obtiene el modelo combinado:

$$y_{it} = \gamma_{00} + \gamma_{10}(\text{día})_{it} + \gamma_{20}(\text{día})_{it}^2 + \gamma_{30}(\text{día})_{it}^3 + u_{1i}(\text{día})_{it} + e_{it}, \quad (3)$$

Los errores de nivel-2 se interpretan como las desviaciones de la pendiente para el país  $i$  con respecto a los valores medios de la población de países. El modelo (3) tiene dos partes, una parte fija  $\gamma_{00} + \gamma_{10}(\text{día})_{it} + \gamma_{20}(\text{día})_{it}^2 + \gamma_{30}(\text{día})_{it}^3$  en la ecuación (3) y una parte aleatoria  $u_{1i}(\text{día})_{it} + e_{it}$ . La parte fija no varía entre los países y la parte aleatoria es susceptible de variar entre los países. Dentro de este contexto los términos  $\gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{30}$  se denominan parámetros fijos y las varianzas/covarianzas de los errores aleatorios se denominan parámetros aleatorios.

En este modelo las tasas de los totales de casos confirmados ( $TTCC$ )<sub>it</sub> por países se describen como una función cúbica del tiempo en el primer nivel. En el segundo nivel se describe la variabilidad entre las curvas de crecimiento de los países. La parte fija del modelo describe la trayectoria de crecimiento promedio a través de tres parámetros: el intercepto promedio  $\gamma_{00}$ , el coeficiente de crecimiento lineal promedio  $\gamma_{10}$ ; el coeficiente de crecimiento cuadrático promedio  $\gamma_{20}$ , y el coeficiente de crecimiento cúbico,  $\gamma_{30}$ . La trayectoria de cada país se describe a través del intercepto específico del país,  $\beta_{0i}$ , el parámetro de crecimiento lineal  $\beta_{1i}$ ; el parámetro de crecimiento cuadrático,  $\beta_{2i}$  y el parámetro de crecimiento cúbico,  $\beta_{3i}$ . Los errores en el nivel-2,  $u_{1i}$ , específicos de los países, se supone que distribuyen normal con media 0 y varianzas  $\sigma_{u_1}^2$ . El término  $e_{it}$  en el nivel-1 es el error en el tiempo  $t$  de la  $i$ -ésima curva. Estos errores se suponen mutuamente independientes y normalmente distribuidos con media 0 y varianza constante  $\sigma_e^2$ .

Durante el desarrollo de la modelación se pueden incluir además, variables explicativas de nivel-2 invariables en el tiempo. Para explicar la variación de las tasas de casos confirmados entre los países, se le adiciona al modelo una covariable dicotómica, que clasifica a los países en dos grupos, atendiendo a la velocidad de crecimiento promedio de sus tasas hasta la fecha de actualización de los datos. De esta manera se construye la variable *Grupo*, indicadora de las siguientes dos

categorías: los países con altas tasas de casos confirmados y los países con bajas tasas de casos confirmados. Esta última categoría se toma como referencia, así, cuando la variable indicadora del grupo al que pertenece el país toma el valor cero, la respuesta en el modelo se interpreta como la tasa promedio para los países pertenecientes a este grupo en el día  $t$ . El modelo cúbico de dos niveles que expresa la relación entre las tasas de casos confirmados y las categorías de la variable *Grupo* se formula de la siguiente manera:

$$y_{it} = \gamma_{00} + \gamma_{10}(\text{día})_{it} + \gamma_{20}(\text{día})_{it}^2 + \gamma_{30}(\text{día})_{it}^3 + \gamma_{11}\text{Grupo}(\text{día})_{it} + u_{1i}(\text{día})_{it} + e_{it}$$

### 1.3 Estimación de los parámetros y prueba de hipótesis

Para estimar los parámetros desconocidos se utilizó el procedimiento Mínimos Cuadrados Generalizados Iterativos Restringidos (RIGLS, por su siglas en inglés), el cual supone que la distribución de los errores es normal multivariada. Para evaluar la bondad de ajuste de cada modelo propuesto se realiza un análisis del cumplimiento de las hipótesis utilizando métodos gráficos de diagnóstico, se identificaron las observaciones atípicas y se analizaron las que tenían una mayor influencia en los modelos. En los casos en que se considere apropiado se excluyen del análisis los países con comportamientos anómalos y se hace un reajuste del modelo. Este proceder iterativo permite identificar subgrupos de países con diferencias marcadas de acuerdo al indicador epidemiológico estudiado.

#### 1.3.1 Curvas predictivas

Para construir las curvas predictivas del indicador epidemiológico  $TTCC$  se ignoran los términos aleatorios de los modelos. La predicción se obtiene evaluando valores específicos de las variables explicativas en los correspondientes modelos marginales (parte fija). Finalmente, se construye el gráfico de las curvas medias estimadas. La trayectoria observada del indicador epidemiológico en Cuba se compara con estas. Finalmente, la estimación de la varianza de los errores de nivel-2 con respecto al término lineal aleatorio puede utilizarse para estimar rangos diarios (intervalos de confianza) del total de casos esperados en Cuba.

## 2. Resultados

En esta sección se muestra un resumen de los resultados de los modelos ajustados para describir el comportamiento de la respuesta media de la  $TTCC$  entre los países seleccionados para el estudio. Los parámetros estimados y sus errores en cada modelo se presentan en la tabla 2.

Para la predicción del  $TCCC$  sólo se consideran las estimaciones de los parámetros fijos; no obstante, en cada modelo debe tenerse en cuenta cuánta de la varianza residual queda como un potencial para ser “explicado” por variables de los dos niveles.



**Tabla 2.** Estimación de los parámetros (y errores estándar) de los modelos A, B, C y D.

| Parámetros                                   | Modelo A         | Modelo B         | Modelo C         | Modelo D         |
|--|------------------|------------------|------------------|------------------|
| <b>Fijos</b>                                 |                  |                  |                  |                  |
| $\gamma_{00}$ ( <i>constante</i> )           | 0,4567 (0,0562)  | 0,4563 (0,0562)  | 0,4228 (0,0649)  | 0,4117 (0,0647)  |
| $\gamma_{10}$ ( <i>día</i> )                 | 0,3115 (0,0113)  | 0,2816 (0,0106)  | 0,2225 (0,0089)  | 0,2322 (0,0086)  |
| $\gamma_{20}$ ( <i>día</i> ) <sup>2</sup>    | -0,0065 (0,0005) | -0,0065 (0,0005) | -0,0036 (0,0003) | -0,0041 (0,0003) |
| $\gamma_{30}$ ( <i>día</i> ) <sup>3</sup>    | 0,0001 (0,0000)  | 0,0001 (0,0000)  | 0,00002 (0,0000) | 0,00003 (0,0000) |
| $\gamma_{11}$ ( <i>Grupo</i> )( <i>día</i> ) |                  | 0,0548 (0,0056)  |                  |                  |
| <b>Aleatorio</b>                             |                  |                  |                  |                  |
| <b>Nivel-2</b>                               |                  |                  |                  |                  |
| $\sigma_{u_1}^2$                             | 0,0010 (0,0003)  | 0,0002 (0,0001)  | 0,0002 (0,0000)  | 0,0001 (0,0000)  |
| <b>Nivel-1</b>                               |                  |                  |                  |                  |
| $\sigma_e^2$                                 | 0,3231 (0,0125)  | 0,3231 (0,0125)  | 0,2704 (0,0126)  | 0,2533 (0,0123)  |
| Log-verosimilitud                            | 2465,3082        | 2419,9174        | 1487,3144        | 1321,9734        |

Para iniciar el proceso de ajustar el efecto del tiempo sobre la tasa del total de casos confirmados se estimaron los parámetros del modelo de tendencia aleatoria o crecimiento incondicional (Modelo A). En este modelo se considera fija la tasa de casos confirmados al comienzo del período evaluado, pero se supone que el ritmo de crecimiento lineal varía entre los países. La estimación de la varianza del coeficiente asociado al término lineal es significativamente diferente de cero, lo que apoya la hipótesis de que el ritmo de crecimiento varía entre los países.

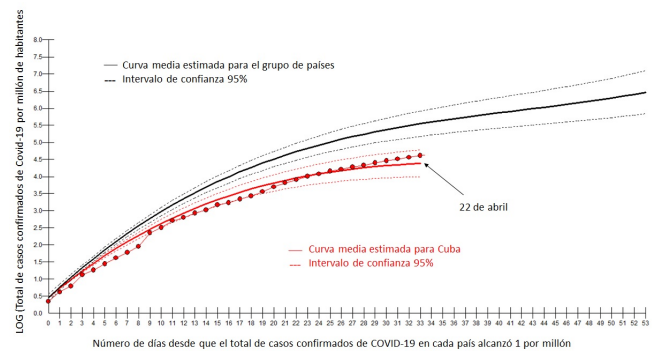
Téngase en cuenta que en el modelo A, la muestra estuvo constituida por 31 países. Este modelo se utilizó para explorar el crecimiento de la *TTCC* en Cuba en relación con el resto de los países hasta la fecha de actualización de los casos.

Las curvas medias de cada uno de los países se construyen considerando los parámetros fijos y los residuos de nivel-2. Para construir la curva media de todos los países considerados en la muestra de estudio solo se consideran las estimaciones de los parámetros fijos.

En la figura 4 se observa que la curva estimada (y la trayectoria en el tiempo de los valores observados) para Cuba difieren significativamente de los valores medios predichos por los parámetros fijos del modelo. Esto es, a medida que el tiempo avanza, el intervalo de confianza de la curva de crecimiento específica de la *TTCC* de Cuba se va alejando de la curva de crecimiento promedio de la *TTCC* del grupo de países con características demográficas similares a Cuba.

En la Figura 5 se presentan los gráficos que muestran los 31 residuos de nivel-2 en orden ascendente, con su correspondiente intervalo de confianza al 95 %. Si se observan los intervalos, se aprecia que la mayoría no contiene al cero. Esto significa que estos países difieren significativamente de los valores promedios predichos por los parámetros fijos. Nótese que Cuba se encuentra aproximadamente en el centro del grupo de países con residuos negativos (crecimiento de la *TTCC* por debajo de la media).

En el modelo B se adiciona la variable de nivel-2, *Grupo*, que explica las diferencias entre los países. La variable *Grupo*

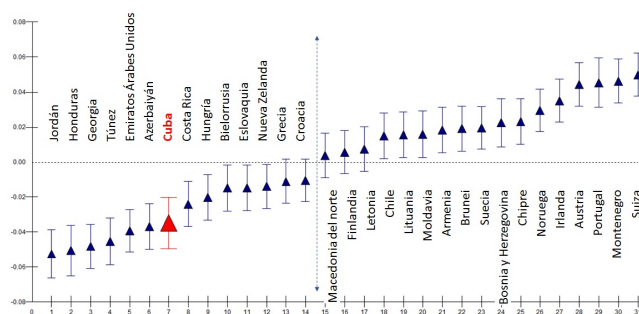
**Figura 4.** Crecimiento de la tasa del total de casos confirmados en Cuba en comparación con el crecimiento promedio (modelo A) del grupo de países con características demográficas similares a Cuba.

se dividió en dos categorías, según si los residuos eran positivos o negativos. El grupo de países con residuos negativos se toma como referencia.

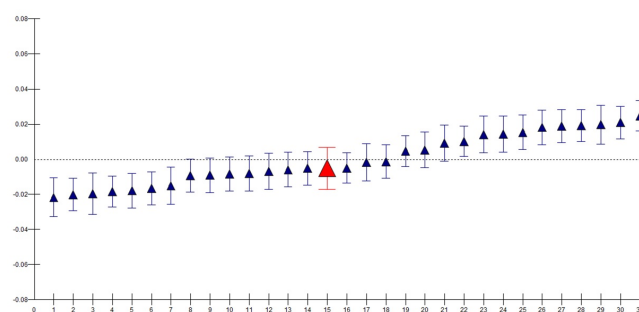
Como se deduce de la tabla 2, existe una fuerte evidencia de la supuesta variación del efecto asociado al grupo, ya que el cambio del modelo (A) al (B) en la log-verosimilitud ( $2465,3082 - 2419,917 = 45,3912$ ), comparado con una distribución  $\chi^2$  es significativo. En la tabla 2 se presentan las estimaciones de los componentes de la varianza de este modelo y es evidente los cambios en los valores de las estimaciones con respecto al modelo A, o sea, la partición en dos grupos de países parece ser un buen predictor de la variabilidad de los parámetros. Correspondientemente, se reduce el valor estimado de la varianza, “entre” de países (de 0,001 a 0,0002).

En la figura 6 se muestran los residuos (asociado al término lineal) y los correspondientes intervalos de confianza al 95 % de los 31 países ordenados en orden ascendente. Después de explicar la variabilidad del crecimiento de las tasas según el grupo al que perteneces el país, Cuba aparece con un país con un crecimiento “típico”.

El grado de cuán apropiadamente el modelo B se ajusta a los datos de los países se expresa por el porcentaje de varianza



**Figura 5.** Residuos de nivel-2 y sus correspondientes intervalos de confianza de los 31 países, según el modelo A.



**Figura 6.** Residuos de nivel-2 y sus correspondientes intervalos de confianza de los 31 países, según el modelo B

explicada en el nivel-2 dada por:

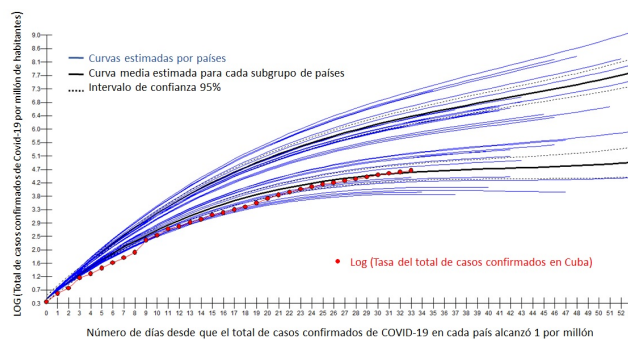
$$\frac{2419,9 [ModeloB]}{(2465,3 [ModeloA] + 2419,9 [ModeloB])} = 0,50$$

Esto significa que aproximadamente el 50 % de la variabilidad entre los países es explicada por la categorización de los países en dos subgrupos, que denominamos: Países con tasas altas y países con tasas bajas.

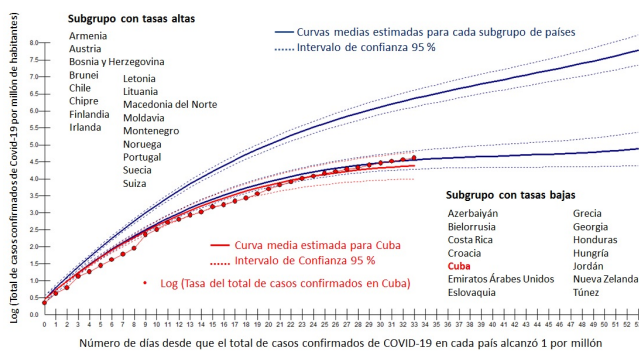
En la figura 7 se muestran las curvas medias de las tasas del total de casos confirmados estimadas para cada uno de los países y su posición relativa con respecto a las dos curvas promedios de cada subgrupo de países. A través de estas curvas se puede apreciar claramente el efecto del tiempo y las diferencias entre los grupos de países de acuerdo a la velocidad de crecimiento de los casos confirmados. En cuanto a la posición del comportamiento de Cuba, queda claro que la trayectoria hasta el día de actualización de los datos, transita muy cerca de la curva promedio de los países con las tasas más bajas.

En la figura 8 se aprecia claramente las diferencias en el comportamiento de la tasa del total de casos confirmados entre los grupos de países. Además, se percibe cómo la curva media de los países en el subgrupo de países con tasas más bajas tiende a un estado de meseta o estabilidad. Se espera que la trayectoria de Cuba transite dentro del intervalo de confianza de esta curva media

La figura 9 muestra gráficamente la comprobación de esta hipótesis. Aquí se evidencia el cumplimiento total de la predicción realizada a partir del modelo B



**Figura 7.** Crecimiento de la tasa del total de casos confirmados en Cuba en comparación con el crecimiento promedio de países con características demográficas similares a Cuba y las curvas medias de dos subgrupos de países (modelo B).

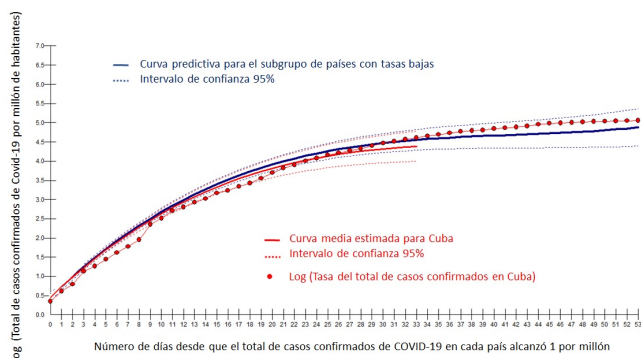


**Figura 8.** Comparación de la trayectoria de la tasa del total de casos confirmados en Cuba con respecto al crecimiento promedio (modelo B) de dos subgrupos de países con características demográficas similares a Cuba

Para comprobar la efectividad y precisión de la estrategia de predicción, se estima el modelo C, donde se utilizan solo los datos para el subgrupo de países con tasas más bajas (14 países), y los datos se actualizan hasta el 7 de mayo de 2020. En la tabla 2, puede observarse que los valores de las estimaciones de la varianza de nivel-2 en el modelo C es significativa, interpretándose que los patrones de crecimiento lineal también difieren entre los países de este grupo. Sin embargo, la varianza estimada de nivel-1, aunque significativa, es más pequeña en comparación con el modelo B. Esto confirma, que independientemente del país, en promedio, el crecimiento diario del total de casos confirmados en este grupo de países es menor que cuando se consideraron todos los países de la muestra.

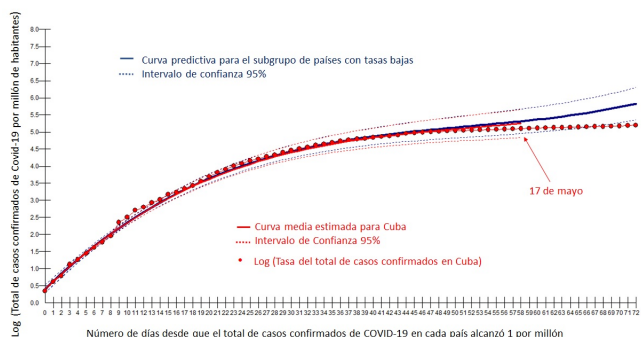
Después de un examen gráfico de los residuos del modelo C, se determina excluir del análisis los datos de Bielorrusia, por considerarse con un comportamiento atípico en relación al resto de los países. Del reajuste se obtiene el modelo D. En la figura 10 se presenta la nueva curva media.

Nótese que los valores observados de las tasas de casos confirmados en Cuba después de la fecha en que se realiza la estimación con datos de 13 países, muestran, como ya



**Figura 9.** Tasa del total de casos confirmados observados hasta 20 días después de realizada la predicción (modelo B).

se esperaba desde la predicción del modelo B, una meseta sostenida, pero ahora traspasa el límite inferior del intervalo de confianza de la curva predictiva según el modelo D. Lo que se interpreta como un escenario favorable.



**Figura 10.** Actualización (modelo D) de la predicción de la tasa del total de confirmados (por millón de habitantes) de COVID-19 en Cuba.

### 3. Conclusiones

Los resultados obtenidos en este trabajo muestran que fue posible, utilizando modelos multinivel, construir curvas predictivas de la tasa del total de casos confirmados de COVID-19 durante la etapa temprana de la epidemia en Cuba. Los modelos propuestos permitieron explicar claramente el comportamiento esperado de este indicador en el país, en comparación con un grupo de países con características demográficas similares a las de Cuba.

Los gráficos construidos a partir de los modelos multinivel predictivos proporcionaron canales de seguimiento que permitieron identificar con claridad las posibles desviaciones en la trayectoria esperada del total de casos confirmados de COVID-19 por millón de habitantes en Cuba. Fue posible, además, calcular los rangos diarios esperados de este indicador epidemiológico.

Suponiendo un contexto en el cual se mantuviera la efectividad de las medidas de control ante la COVID-19 en Cuba,

los modelos brindaron un escenario de confianza en el que se podría esperar que los valores del indicador epidemiológico en el tiempo no excederían significativamente el límite superior de los rangos diarios estimados alrededor de la curva media de un subgrupo de países con las tasas más bajas entre el grupo de países con características demográficas similares a Cuba.

Las curvas predictivas, resultantes de los modelos propuestos, se construyeron tratando de mantener un diseño simple que convierte la estrategia en una herramienta prácticamente útil; no obstante, la flexibilidad del enfoque multinivel permite la construcción de modelos alternativos considerando otros indicadores epidemiológicos y la adición de variables contextuales capaces de explicar las diferencias entre los países.

### Agradecimientos

Agradezco a todos los compañeros que durante las reuniones del equipo multidisciplinario dedicado al estudio y modelación de la epidemia de COVID-19 en Cuba me aportaron información relevante para el desarrollo de este trabajo. En particular me gustaría reconocer a Yudivián Almeida y a Gabriel Gil. Igualmente agradezco el apoyo del Instituto de Cibernética, Matemática y Física (ICIMAF), en especial a Victoria Hernández, por sus comentarios y útiles sugerencias, así como por la revisión editorial de este artículo.

### Referencias

- [1] Almeida Cruz Y., comunicación personal, Universidad de La Habana, 20 de abril de 2020.
- [2] Giordano G., Blanchini F., Bruno R., Colaneri P., Filipo A., Matteo A., Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med* 26, 855–860, 2020. <https://doi.org/10.1038/s41591-020-0883-7>.
- [3] <https://github.com/owid/covid-19-data>.
- [4] Goldstein H. *Multilevel Statistical Models*. 2nd. Ed. Halsted Press, New York, 1995.
- [5] Holmdahl I. y Buckee C. Wrong but Useful - What COVID-19 Epidemiologic Models Can and Cannot Tell Us. *N Engl J Med*, 383, 303–305, 2020.
- [6] Jewell N. P., Lewnard J. A. y Jewell B. L. Caution warranted: using the Institute for Health Metrics and Evaluation model for predicting the course of the COVID-19 pandemic. *Ann Intern Med*, 2020.
- [7] Kermack W. O. y McKendrick A. G. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, Vol. 115, No. 772, 700–721, 1927.

- [8] Raudenbush S. W. y Bryk A. S. Hierarchical linear models: Applications and data analysis methods. (2nd Edition). Thousand Oaks, CA: Sage, 2002.
- [9] Roser M., Ritchie H, Ortiz-Ospina E. y Hasell J. Coronavirus Pandemic (COVID-19). Published online at OurWorldInData.org. Retrieved from: <https://ourworldindata.org/coronavirus> [Online Resource], 2020.
- [10] Singer J.D. y Willett J.B. Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford University Press, 2003.
- [11] Snijder T. A. B., y Bosker R. J. Multilevel analysis: Introduction to basic and advanced multilevel modelling. London: Sage., 1999.
- [12] Verbeke G. y Molenberghs G. Linear Mixed Models for Longitudinal Data Springer Verlag, New York, 2000.
- [13] World Health Organization (WHO): <https://www.who.int/>. Accessed 12 Apr 2020.