

# La dinámica de transmisión del covid-19 desde una perspectiva matemática

**Raúl Isea**

Fundación Instituto de Estudios Avanzados (IDEA)  
ORCID: 0000-0002-6318-3428  
raul.isea@gmail.com  
Venezuela

**Fecha de recepción: 24 - 03 - 2020    Fecha de aceptación: 17- 04- 2020**

## Resumen

El presente trabajo plantea un modelo matemático capaz de reproducir la dinámica de transmisión del nuevo coronavirus (Covid-19) en grupos humanos a partir de un sistema de ecuaciones diferenciales. Para ello se dividió a la población en tres clases diferentes dependiendo de los estadios de la enfermedad: aquellos que son susceptibles de contraer el virus, los infectados y los recuperados. Los parámetros fueron

determinados por mínimos cuadrados a partir de los registros diarios del Covid-19 realizados por la Universidad Johns Hopkins, y se validó en cuatro países seleccionados al azar: China, Estados Unidos, Brasil y Venezuela. Simultáneamente, se exploró la calidad de los datos para detectar cualquier manipulación o alteración de las cifras de contagios en estos cuatro países, a partir de dos metodologías computacionales empleadas en análisis forense de información digital. Así, será po-

sible predecir *grosso modo* el número de contagios en el tiempo: a modo de ejemplo, se estimó que pudieran darse 597 casos de contagios en la República Bolivariana de Venezuela hasta el 22 de junio del presente año en función de la información analizada hasta el 20 de abril, fuertemente influenciada por los brotes detectados.

**Palabras clave:** covid-19; pandemia; ecuaciones diferenciales; modelo; ley de Benford

# The Transmission Dynamics of Covid-19 from a Mathematical Perspective

## Abstract

The present work proposes a Mathematical model capable of reproducing the transmission dynamics of the new coronavirus (Covid-19) in human groups, from a system of differential equations. The total population was divided into three different types: susceptible, infected, and recovered. Parameters were developed using the least

squared method, based on Johns Hopkins' Covid-19 data, and were validated in four countries: China, the United States, Brazil and Venezuela. Simultaneously, the quality of the data was explored to detect any manipulation or alteration of the numbers of infections in these four countries, based on two computational methodologies used in forensic analysis of digital information. Finally, it will be possible to estimate

that 597 cases of infection could occur in the Bolivarian Republic of Venezuela, until June 22 of this year, based on the information analyzed up to April 20, heavily influenced by the detected break out.

**Key words:** Covid-19; pandemic; model; outbreak; Benford law

## Introducción

El nuevo coronavirus (Covid- 19) fue declarado una pandemia por la Organización Mundial de la Salud el 11 de marzo de 2020. Los primeros brotes comenzaron a registrarse en el mercado de Wuhan, en China, y hasta el 24 de abril de 2020 se cuentan más de dos millones setecientos mil infectados, con al menos ciento noventa mil muertos distribuidos en 180 países en todo el mundo. Lamentablemente, no existe una vacuna para combatir esta enfermedad, por lo que la prevención basada en medidas de cuarenta social han sido los instrumentos que están ayudando a contrarrestar la propagación del virus. Por esa razón, es necesario desarrollar modelos matemáticos que permitan monitorear y predecir la dinámica de transmisión de contagio causado por este nuevo coronavirus.

Al revisar la literatura científica, se encuentra una amplia gama de modelos matemáticos basados en ecuaciones diferenciales (p. e., Feng, 2007; Huang et al, 2020), sin embargo, muchos de ellos no reproducen un crecimiento del tipo logístico de los casos infectados. Por otra parte, se necesita verificar la calidad de los datos aportados por cada país, y por esa razón se realiza una exploración empleando dos técnicas tradicionales en análisis de datos digitales, que permiten detectar cualquier manipulación.

Este trabajo plantea entonces un modelo matemático basado en tres ecuaciones diferenciales atendiendo a los diferentes estadios de la enfermedad, y simultáneamente se hace énfasis en la necesidad de verificar la confiabilidad de los datos.

## El modelo propuesto

Se propone un modelo del tipo determinístico basado en uno que reproduce un crecimiento del tipo logístico del número de casos infectados. Para ello se dividió a la población total de acuerdo a tres estadios, es decir: susceptibles a contraer la enfermedad (S), infectados por el coronavirus (I), y, finalmente, las personas que se recuperan de la enfermedad (R). Es importante destacar que se considera una población del tipo cerrada, esto quiere decir que no se incluyen nacimientos ni efectos por casos de contagios por inmigración en el país. Por otra parte, se asume que las personas recuperadas son inmunes de por vida, y por tanto incapaces de propagar la enfermedad a otras personas. Teniendo en cuenta estas consideraciones, el sistema de ecuaciones es el siguiente:

$$\begin{aligned} \frac{dS(t)}{dt} &= - \frac{\beta(t) \cdot S(t) \cdot I(t)}{N} - \alpha S(t) \\ \frac{dI(t)}{dt} &= \frac{\beta(t) \cdot S(t) \cdot I(t)}{N} - (\alpha + \gamma) \cdot I(t) \\ \frac{dR(t)}{dt} &= \gamma \cdot I(t) - \alpha \cdot R(t) \end{aligned}$$

donde N es el número total de habitantes. B (t) es una variable que depende del tiempo capaz de reproducir un crecimiento logístico de los casos infectados, igual a  $\beta([\mu+I(t)])$ . Por otra parte  $\alpha$ , representa la tasa de muertes, mientras que toma en cuenta la proporción de personas recuperadas.

Por tanto, el modelo requiere determinar cuatro constantes diferentes ( $\beta$ ,  $\alpha$ ,  $\gamma$ ,  $\mu$ ), que permitirían establecer la dinámica de transmisión de contagios en un determinado país. Para ello se realiza un ajuste por mínimos cuadrados con los datos registrados por Covid-19 que están disponibles en la Universidad Jo-

hns Hopkins. Por esta razón, y a juicio del autor, es necesario comenzar a explorar la calidad de los datos registrados para detectar cualquier anomalía o manipulación en ellos que pueda influir en los resultados obtenidos en el trabajo.

## Validación de los datos de personas infectadas

Existe una amplia gama de trabajos científicos donde se han empleado técnicas computacionales en análisis forense de datos para detectar cualquier irregularidad, las cuales están basadas en la ley de Benford (Fewster, 2009). Dicha metodología se ha validado en diversos escenarios, como por ejemplo, en la detección de fraudes en declaraciones de impuesto, alteración en registros bancarios, o ratificar datos en genómica, entre otros.

Recientemente, se publicó un trabajo (Zhang, 2020) donde se analiza la confiabilidad de las cifras sobre los casos infectados en China hasta el 10 de febrero de 2020, basada justamente en la ley de Benford. El investigador concluye que no hay manejo extraño de la información a raíz de que el valor estadístico p-valor (*p value*) es del 92.8%.

Por ello, se extiende y se complementa el análisis realizado por Zhang a todos los datos registrados hasta el 20 de abril de 2020 en cuatro países, mediante un estudio de frecuencia del primer dígito de acuerdo a la ley de Benford; y además, se implementa una segunda prueba computacional llamada prueba de arco de mantisa (del inglés *Mantissa Arc Test*) obtenida tras el estudio del segundo dígito (Alexander, 2009).

## Desarrollo

Los datos registrados de la pandemia de coronavirus fueron obtenidos de la base de datos mantenida por la Universidad Johns Hopkins, la cual se consul-

ta en línea y se actualiza todos los días (disponible en <http://coronavirus.jhu.edu>). De ella se extrajeron los casos infectados, recuperados y fallecidos de cuatro países: China, Estados Unidos, Brasil y Venezuela.

El próximo paso fue examinar la calidad de los datos de acuerdo a dos pruebas computacionales basadas en la frecuencia de aparición del primer dígito, así como la prueba de arco de mantisa del segundo dígito para descartar cualquier alteración o manipulación. Con ese fin se diseñó un programa en el lenguaje de software libre Python (versión 3.5) con el que se determinan la frecuencia de aparición del primer dígito de izquierda a derecha, y posteriormente, la probabilidad (*Prob*), de acuerdo a la siguiente ecuación:

$$Prob(d_i) = \log_{10}\left(1 + \frac{1}{d_i}\right)$$

donde *i* corresponde a los valores que van desde 1 hasta 9 (Forman, 2010). Asimismo, y para evitar un sesgo en el resultado obtenido con el primer método, se implementó otro algoritmo que permite realizar la prueba de arco de mantisa basada en la determinación del centro de masa del conjunto de valores obtenidos de la mantisa cuando se considera un círculo unitario (Forman, 2010).

El próximo paso fue realizar un ajuste por mínimos cuadrados de los casos infectados con los datos que va generando el modelo, según la metodología que ha sido validada en diversas publicaciones científicas (p.e., Isea y Lonngren, 2015; Isea, 2016; Isea y Lonngren, 2017; Isea, 2018; Isea, 2019; Isea y Lonngren, 2020).

Finalmente, se resuelve numéricamente el sistema de ecuaciones empleando el método de Runge-Kutta de 4to orden (en vista de que se minimiza el error al utilizar polinomios de Taylor de hasta cuarto orden), gracias a los parámetros obtenidos del ajuste anterior, como se explica en la próxima sección.

## Resultados

En la Tabla 1 se presentan los resultados obtenidos tras examinar y monitorear la integridad de los datos de casos infectados en China, Estados Unidos, Brasil y Venezuela. Como se aprecia, la primera prueba revela que China presenta un excelente acuerdo y por ende, no hay evidencia de manipulación en sus datos. Por otra parte, y quizás porque aún existen pocos registros en Estados Unidos, este país no cumple la primera prueba basada en la frecuencia del primer dígito (se resaltó en rojo la cifra), mientras que Brasil y Venezuela superan dicho umbral, pero estos tres países (Estados Unidos, Brasil y Venezuela) presentan un valor alto en la dispersión de sus datos, como se lee en la columna de desviación cuadrática media (Tabla 1).

La segunda metodología, que sigue la prueba de arco de mantisa a partir del segundo dígito, revela que Estados Unidos, Brasil y Venezuela no alteran sus datos, aunque superan el valor de L2 cuando se comparan con el resultado obtenido en China (tabla 1). Sin embargo, es necesario realizar mayores estudios con los datos de Estados Unidos y del resto de los países para garantizar su integridad.

**Tabla 1. Resultados obtenidos con dos pruebas computacionales basadas en cálculos con el primero y segundo dígito de acuerdo a la ley de Benford, tras un análisis de los casos infectados hasta el 20 de abril de 2020.**

País	Primer dígito		Segundo dígito		
	p-value	Desviación cuadrática media	Prueba de arco de mantisa	$L^2$	Desviación cuadrática media
China	0.9311	0.0167	0.6856	0.0038	0.0700
Estados Unidos	<b>0.0159</b>	0.0481	0.1150	0.0313	0.0117
Brasil	0.3939	0.0384	0.1256	0.0432	0.0113
Venezuela	0.1388	0.0651	0.3054	0.0475	0.0159

De los resultados obtenidos anteriormente pudiera inferirse que no existe ninguna manipulación, y por tanto se procedió a realizar el ajuste por mínimos cuadrados para determinar los parámetros que permitirán describir la

dinámica de contagio en cada país. En tal sentido, se resumen en la Tabla 2 los datos y los resultados conseguidos. En vista de que hay más incógnitas que ecuaciones, se consideró que el valor de  $\beta$  es el mismo para todos los países,

es igual a 0.00003914. De ese modo, se logró estimar que el valor de  $\mu$  es  $0.429 \pm 0.068$ , cuyo margen de error permite ajustarlo para todos los países. Este último resultado debe analizarse con más detenimiento en próximos trabajos.

**Tabla 2. Datos y resultados obtenidos tras la evaluación por mínimos cuadrados del sistema de ecuaciones**

País	Población	Fecha inicial considerada en el modelo	Número de casos considerados	$\beta/\mu$
China	1.439.323.774	31/diciembre/2019	113	0.638
Estados Unidos	331.002.647	23/enero/2020	90	0.610
Brasil	212.559.409	26/febrero/2020	53	0.689
Venezuela	28.435.943	15/marzo/2020	38	0.849

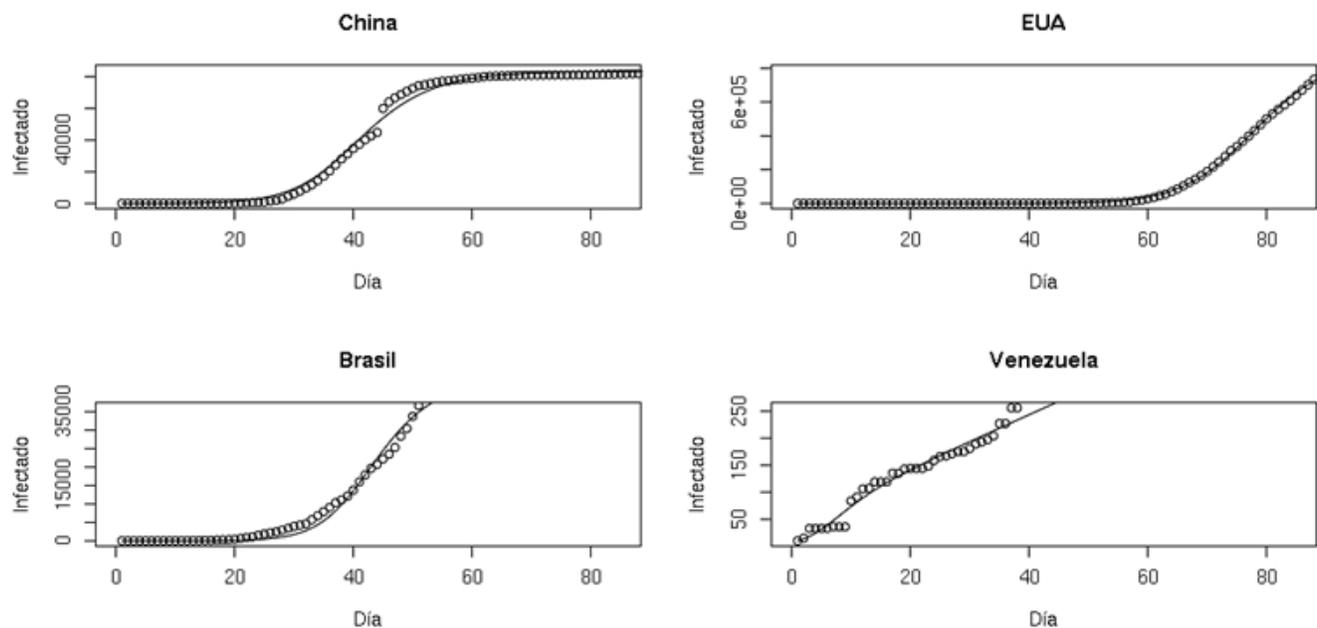
Gracias a estos valores, los dos parámetros restantes se presentan en la última columna de la Tabla 2. Probablemente el resultado tan alto obtenido en Venezuela con respecto a los otros países se deba al poco número de casos registrados en el modelo.

En la Figura 1 se muestran los resultados tras la integración numérica del

sistema de ecuaciones que registran los casos infectados en cada país. Dicha integración fue realizada con el método de Runge-Kutta de 4to orden empleando para ello los parámetros derivados del ajuste por mínimos cuadrados, donde los datos registrados por la Universidad Johns Hopkins se representan en un círculo, mientras que con una línea se indica el resultado de la integración

numérica. Para visualizar mejor los resultados, el tiempo se fijó en cero en la figura, pero se debe tener en cuenta que cada valor inicial corresponde al 31 / diciembre / 2019, 23 / enero / 2020, 26 / febrero / 2020 y 15 / marzo / 2020 para China, Estados Unidos, Brasil y Venezuela, respectivamente.

**Figura 1. Inintegración numérica del sistema de ecuaciones que registran los casos infectados en cada país**



En la Figura 1 se presentan los resultados obtenidos del ajuste por mínimos cuadrados de los distintos casos infectados en China, Estados Unidos, Brasil y Venezuela. En un círculo se muestran los casos registrados por la Universidad Johns Hopkins y en una línea continua, el resultado de la integración numérica obtenido del sistema de ecuaciones. Por claridad en los resultados, el tiempo en cada gráfica comenzó en cero, pero corresponde al 31/diciembre/2019, 23/

enero/2020, 26/febrero/2020 y 15/marzo/2020 para estos países.

La ventaja de emplear este tipo de análisis consiste en que sería posible predecir el número de contagios en un determinado tiempo, con base exclusivamente en los datos registrados. Con los que se han recopilado en Venezuela hasta ahora, sería una aventura predecirlo, pero respetando la metodología antes descrita, se pronostican 597 casos de

contagios hasta el 22 de junio de 2020. Dicho número está basado en los brotes locales recientemente registrados en el país, sin tener en cuenta factores que pueden elevar esa cifra por efectos de la inmigración. De más está mencionar que los brotes recientes en el país han sido originados por el incumplimiento de la cuarentena social, de modo que si se respeta, la cifra debe disminuir.

## Conclusiones

El presente trabajo propone un modelo matemático que puede ayudar a describir la dinámica de transmisión del Covid-19 empleando un sistema de tres ecuaciones diferenciales. Para lograrlo, se utilizaron los datos registrados por la Universidad Johns Hopkins en su base de datos hasta el 20 de abril de 2020, considerando únicamente cuatro países: China, Estados Unidos, Brasil y Venezuela. De acuerdo a una inspección de los datos publicados por esa universidad empleando dos técnicas computacionales aplicadas en estudios forenses de información, China, Brasil y Venezuela superan las dos pruebas, mientras que Estados Unidos solo uno de los dos test realizados. Este resultado quizás pueda estar influenciado por el continuo incremento de casos diarios, y en unas semanas será posible verificarlo. La principal ventaja del modelo es su carácter predictivo; a modo de ejemplo, se estimó que en Venezuela pueden llegar a registrarse 597 casos hasta el 22 de junio de 2020 de acuerdo a los datos disponibles hasta el 20 de abril. Lógicamente, dicho número debería disminuir si se cumple la cuarenta social y se controlan los brotes que pudieran llegar a ocurrir en la frontera del país.

## Referencias Bibliográficas

- Alexander, J. C. (2009). Remarks on the use of Benford's Law. En Red. Disponible en SSRN, 1505147.
- Feng, Z. (2007). Final and peak epidemic sizes for SEIR models with quarantine and isolate. *Mathematical Biosciences and Engineering*, 4(4): 675-686.
- Fewster, R. M. (2009). A simple explanation of Benford's Law. *The American Statistician*, 63(1): 26-32.
- Forman, A. K. (2010). The Newcomb-Benford law in its relation to some common distributions. *PLoS ONE*, 5, e10541.
- Huang, Y.; Yang, L.; Dai, H.; Tian, F. & Chen, K. (2020). Epidemic situation and forecasting of COVID-19 in and outside China. DOI: Recuperado en: <http://dx.doi.org/10.2471/BLT.20.251561>.
- Isea, R. (2016). A Preliminary Mathematical Model for the Dynamic Transmission of Dengue, Chikungunya and Zika. *American Journal of Modern Physics and Application*, 3(2): 11.
- Isea, R. (2018). Analytical solutions for the initial steps of the intracellular dynamics of influenza virus. *Acta Scientific Microbiology*, 1(6): 6-8.
- Isea, R. (2019). Modelando la relación de pareja con ayuda de ecuaciones diferenciales. *Revista Electrónica, Conocimiento Libre y Licenciamento*, 10(19): 94-98.
- Isea, R. & Lonngren, K. E. (2015). A Mathematical Model of Cancer under Radiotherapy. *International J. of Public Health Research*, 5(3): 340.
- Isea, R. y Lonngren, K. E. (2017). A New Variant of the SEIZ Model to Describe the Spreading of a Rumor. *International Journal of Data Science and Analysis*, 3(4): 28-31.
- Isea, R. y Lonngren, K. E. (en prensa). A mathematical model that could explain the social dynamics in a country. *Ciencia en Revolución*, 6(17).
- Zhang, J. (2020). *Testing Case Number of Coronavirus Disease 2019 in China with Newcomb-Benford Law*. Extraído el 13 febrero, 2020 de arXiv: 2002.05696.