

Evaluación preliminar de un sistema de vigilancia en salud pública usando la ley de Benford

Preliminary evaluation of a public health surveillance system using the Benford's law

Jorge Homero Wilches Visbal^{1*} <https://orcid.org/0000-0003-3649-5079>

Midian Clara Castillo Pedraza¹ <https://orcid.org/0000-0003-3170-3959>

Oskarly Pérez Anaya¹ <https://orcid.org/0000-0002-0701-7847>

¹Universidad del Magdalena, Facultad de Ciencias de la Salud. Santa Marta, Colombia.

*Autor para la correspondencia: jhwilchev@gmail.com

RESUMEN

La COVID-19 ha significado un desafío sin precedentes para los sistemas de vigilancia en salud pública de Colombia y el mundo. La calidad del informe de los datos es importante para la elaboración de medidas de contención apropiadas. Como el informe de datos no es ajeno a manipulaciones o deficiencias del sistema de vigilancia, técnicas que ayuden a detectar tales anomalías podrían contribuir a un desempeño eficiente de las medidas. En este trabajo se usó la ley de Benford para analizar preliminarmente el desempeño del sistema de vigilancia en salud pública de Santa Marta (Colombia) a partir del informe de los fallecidos en los dos primeros meses de pandemia. Los resultados mostraron serias discrepancias en el número de fallecidos respecto a lo esperado por la ley de Benford. Una investigación más extensa y exhaustiva permitiría una evaluación más precisa del sistema de vigilancia en salud pública de la ciudad.

Palabras clave: ley de Benford; vigilancia en salud pública; Santa Marta; COVID-19; primer dígito; fallecidos.

ABSTRACT

COVID-19 has represented an unprecedented challenge for the public health surveillance systems in Colombia and the world. The quality of the data reporting is important for the development of appropriate containment measures. As reported data are not immune to manipulations or deficiencies of the surveillance system, techniques that help to detect such anomalies could make the measures more efficient. In this work, Benford's law was used to preliminarily analyse the performance of public health surveillance system in Santa Marta (Colombia) based on the report of deaths in the first two months of the pandemic. Results showed serious discrepancies from what was expected by Benford's law. A more thorough and extensive investigation could allow a more precise and comprehensive examination of the city's public health surveillance system.

Keywords: Benford's law; public health surveillance; Santa Marta; COVID-19; first digit; deceased.

Recibido: 20/12/2020

Aceptado: 20/01/2021

Introducción

La pandemia por COVID-19, aún en curso, ha afectado a más de 70 millones de seres humanos alrededor en el mundo, de los cuales más 1,6 millones han fallecido hasta diciembre de 2020. En Colombia, en esa fecha, se infectaron 1,4 millones, de los que han fallecido 40 000.⁽¹⁾ A diferencia de las epidemias causadas por el SARS-CoV-1 y el MERS-CoV, la originada por el SARS-CoV-2 ha recibido una atención mediática, académica y gubernamental sin precedentes debido a su alcance e impacto global.⁽²⁾

Dado que para la elaboración de estrategias de contención del coronavirus es crucial el informe oportuno y preciso de nuevos casos y fallecidos, el cubrimiento global del progreso de la pandemia ha prestado un gran servicio.^(2,3) No obstante, como los datos informados no están exentos de errores, descuidos o manipulaciones,^(3,4) principios matemáticos que adviertan tales desaciertos son de especial utilidad.

Un principio matemático relevante para la detección de registros anómalos es la denominada ley de Benford o ley del primer dígito. Esta ley surgió del hecho de que diversos conjuntos de datos de la vida real no se distribuyen uniformemente,

sino que están sesgados hacia el menor dígito. Así, la ley de Benford establece que, en un conjunto determinado de números, aquellos con primer dígito igual a 1, son más frecuentes que los que empiezan por el subsiguiente y así sucesivamente. Por tanto, la probabilidad de que n sea el primer dígito de un número cualquiera, se calcula como:⁽⁵⁾

$$P(n) = \log\left(\frac{n+1}{n}\right)$$

donde:

$$n \in \{1,2,3, \dots, 9\}$$

Esta ley ha sido exitosamente empleada para la detección de fraudes fiscales o electorales, así como en física⁽⁵⁾ y epidemiología.^(2,4,5,6) Ya que las enfermedades infecciosas siguen un patrón exponencial, la ley de Benford es particularmente útil para estudiar su comportamiento, al menos, en sus primeros estadios.^(2,4) Cabe anotar que esta la ley es invariable a la escala y se aplica a conjuntos grandes de números no aleatorios.⁽⁷⁾

Siendo así, la ley de Benford puede ser utilizada para evaluar el desempeño de sistemas de vigilancia en salud pública en el monitoreo de enfermedades infecciosas potencialmente pandémicas.^(4,8,9) Un buen sistema de vigilancia en salud pública es aquel que logra retrasar el crecimiento exponencial de la curva epidémica o, en su defecto, aplanarla a fin de distribuir más espaciadamente en el tiempo a los pacientes contagiados para con ello evitar el colapso de los hospitales y el sistema de salud.⁽⁴⁾

Con esto, la ley de Benford puede usarse para medir el grado de éxito de las intervenciones de control epidemiológico, pues cuando la curva de crecimiento de una epidemia no siga una distribución exponencial, se habrá logrado el aplanamiento de la curva epidémica, reflejado en que el número de infectados y/o fallecidos no obedecerá a tal ley.⁽⁴⁾ Asimismo, la ley de Benford puede emplearse para evaluar la calidad de los datos de infectados y fallecidos informados por las entidades territoriales en los primeros días de epidemia, de manera que si tales datos no siguen la ley, se dirá que las entidades poseen fallas en sus sistemas de vigilancia en salud pública.⁽⁹⁾

En este trabajo se aplicó la ley de Benford para analizar el patrón de los individuos fallecidos durante los dos primeros meses de curso de la pandemia en la ciudad de Santa Marta (Colombia), con el propósito de detectar fallas en el sistema de vigilancia salud pública local y tratar de explicar sus razones. Para la obtención de los resultados y gráficas se utilizó el paquete *benford.analysis*⁽¹⁰⁾ bajo la plataforma computacional RStudio. Los resultados arrojaron serios desvíos de los datos informados con respecto a los esperados por la ley de Benford.

Desarrollo

Se aplicó la ley de Benford para analizar el patrón de los fallecidos por COVID-19 en la ciudad colombiana de Santa Marta, entre el 20 de marzo y el 20 de mayo de 2020. El número de fallecidos por día fue extraído de la base de datos del Instituto Nacional de Salud (INS) de Colombia⁽¹¹⁾. Tanto el número de fallecidos como el conteo del primer dígito se visualizan en la tabla 1.

Tabla 1 - Número de fallecidos informados por día y el primer dígito asociado, durante los dos primeros meses de pandemia

Fecha	Fallecidos	Primer dígito
20/03/2020	1	1
21/03/2020	1	1
22/03/2020	1	1
23/03/2020	1	1
24/03/2020	1	1
25/03/2020	1	1
26/03/2020	1	1
27/03/2020	1	1
28/03/2020	1	1
29/03/2020	1	1
30/03/2020	1	1
31/03/2020	1	1
01/04/2020	1	1
02/04/2020	1	1
03/04/2020	1	1
04/04/2020	1	1
05/04/2020	1	1
06/04/2020	1	1
07/04/2020	1	1
08/04/2020	3	3
09/04/2020	5	5
10/04/2020	7	7
11/04/2020	10	1
12/04/2020	10	1
13/04/2020	10	1
14/04/2020	10	1
15/04/2020	10	1
16/04/2020	11	1
17/04/2020	11	1
18/04/2020	11	1
19/04/2020	11	1
20/04/2020	12	1
21/04/2020	12	1
22/04/2020	13	1
23/04/2020	16	1
24/04/2020	17	1

25/04/2020	17	1
26/04/2020	17	1
27/04/2020	17	1
28/04/2020	17	1
29/04/2020	17	1
30/04/2020	18	1
01/05/2020	18	1
02/05/2020	18	1
03/05/2020	18	1
04/05/2020	19	1
05/05/2020	19	1
06/05/2020	19	1
07/05/2020	19	1
08/05/2020	19	1
09/05/2020	19	1
10/05/2020	19	1
11/05/2020	19	1
12/05/2020	19	1
13/05/2020	19	1
14/05/2020	20	2
15/05/2020	20	2
16/05/2020	20	2
17/05/2020	20	2
18/05/2020	20	2
19/05/2020	20	2
20/05/2020	20	2

A partir de los datos de la tabla 1, se puede realizar una comparación entre el primer dígito de los fallecidos informados por día y los esperados de acuerdo con la ley de Benford, como se muestra en la figura.

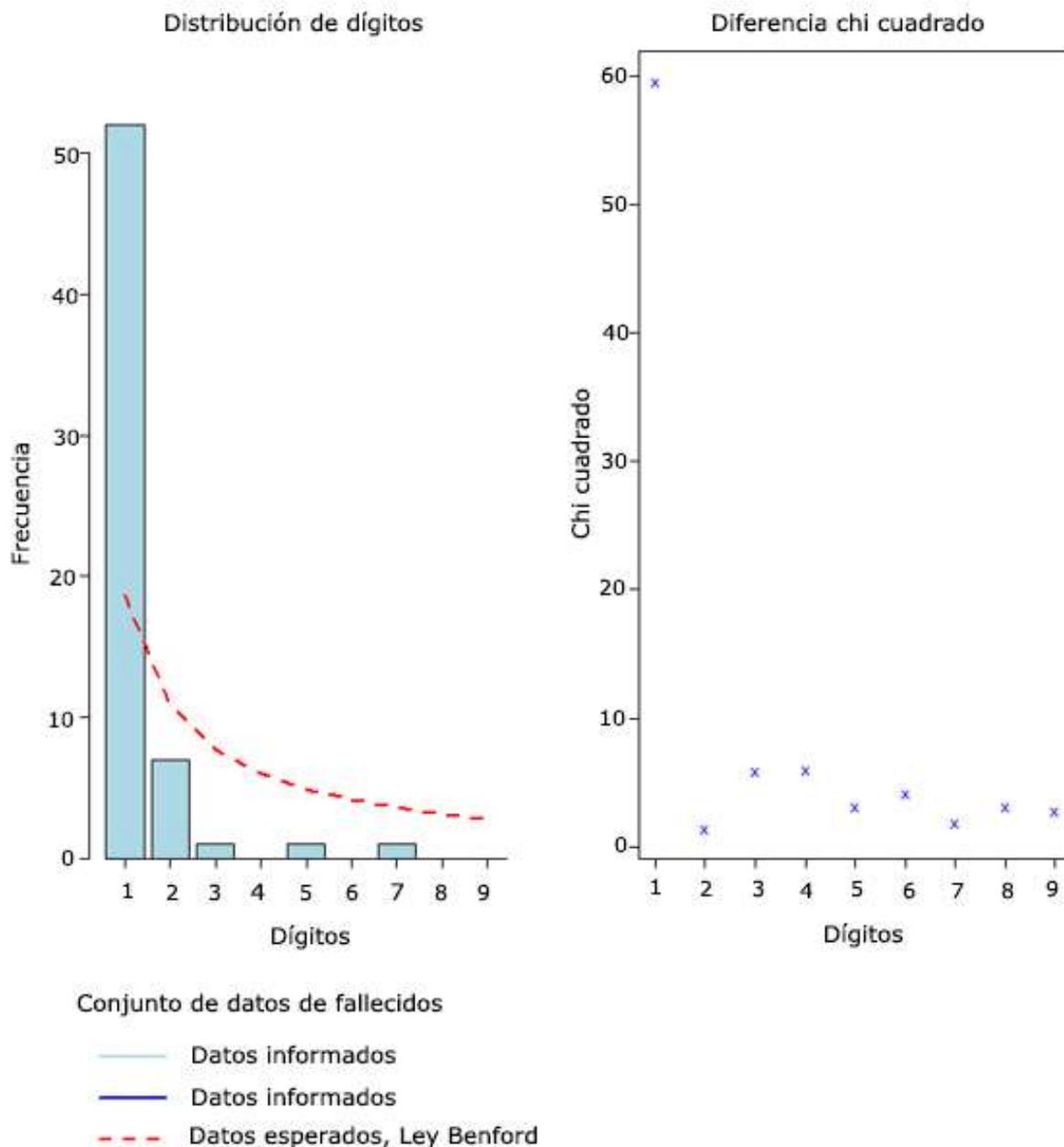


Fig. - Frecuencia del primer dígito de datos informados (barras azules) contra esperados por la ley de Benford (línea roja punteada), y diferencia de la prueba chi cuadrado entre datos informados y esperados.

En la figura se observa una sobrestimación del primer dígito 1 y una subestimación para los demás primeros dígitos con respecto a lo esperado por la ley de Benford. La prueba chi cuadrado indica que los datos con primer dígito 1, en especial, presentan grandes desvíos.

En la tabla 2 se encuentran las variables usadas para evaluar si los desvíos son estadísticamente significativos, en cuyo caso, señalaría que los datos informados merecerían ser investigados a fin de dilucidar tales discrepancias.

Tabla 2 - Variables de análisis estadístico de la Ley de Benford

Variabes de análisis	Valor calculado
Media de la Mantis	0,161
Chi cuadrado	87,978
Desviación absoluta media	0,119
Factor de distorsión	-58,39

DAM: desviación absoluta media, FD: factor de distorsión.

Según *Nigrini*,⁽¹²⁾ si los datos informados siguieran la ley de Benford (hipótesis nula), entonces, la media de la Mantis debería ser 0,5; la prueba chi cuadrado tendría valores cercanos a 0; la DAM no debería ser superior a 0,015 y el factor de distorsión debería ser nulo (si es negativo, se dice que los números están infravalorados; si es positivo, sobrevalorados).

A partir de los datos de la tabla 2 se observa que ninguna de las condiciones anteriores se cumple. Para confirmar esto, se encontró que el valor-*p* de la prueba chi cuadrado fue de $1,976 \times 10^{-15}$. Sin embargo, como la prueba chi cuadrado es muy sensible al tamaño de la muestra, es preferible usar la prueba DAM, pues es independiente del número de registros.^(2,12)

Es posible que las discrepancias encontradas en el periodo de análisis sean la insuficiente infraestructura tecnológica para el procesamiento e informe oportuno de datos; la falta de descentralización de las decisiones; el retraso en la capacitación del personal y coordinación efectiva con los entes estatales de salud y la tardanza en el informe toda vez que el procesamiento se hacía en Bogotá o Barranquilla.⁽⁸⁾ De hecho, a partir del 13 de mayo, se comenzaron a procesar pruebas PCR en Santa Marta.⁽¹³⁾

La COVID-19 ha dejado unas enseñanzas en relación con la importancia de la calidad de los datos informados en una epidemia: i) como las políticas públicas se basan en la evidencia, la información entregada por los entes territoriales debe ser transparente y divulgada casi que en tiempo real, de modo que se puedan evaluar y corregir a tiempo las intervenciones en el sistema de vigilancia en salud pública; ii) incorporación de tecnologías más sofisticadas para la recopilación, procesamiento y emisión de los datos; iii) disponibilidad de datos abiertos por departamento para descentralizar la toma de decisiones.^(2,8)

Conclusiones

Sistemas de vigilancia en salud pública pueden ser efectivamente analizados haciendo uso de la ley de Benford.

En este trabajo se muestra que, en los primeros dos meses de pandemia, el informe de los fallecidos en Santa Marta presentó serias discrepancias respecto a lo esperado por la ley de Benford, lo cual parecería indicar que el sistema de

vigilancia en salud pública de la ciudad no tuvo un buen desempeño en ese periodo.

Una investigación más exhaustiva y extensa podría permitir un examen más preciso y general del sistema de vigilancia en salud pública de la ciudad. Como perspectiva, podría aplicarse la ley de Benford para analizar el éxito de las medidas de contención en la ciudad, involucrando los casos acumulados y una ventana de tiempo mayor.

Referencias bibliográficas

1. European Centre for Disease Prevention and Control. COVID-19 situation update worldwide. 2020 [access: 19/12/2020];1. Available from: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>
2. Silva L, Figueiredo Filho D. Using Benford's law to assess the quality of COVID-19 register data in Brazil. *J Public Health (Bangkok)*. 2020 [access: 19/12/2020];1-4. Available from: <https://academic.oup.com/jpubhealth/advance-article/doi/10.1093/pubmed/fdaa193/5937152>
3. Sambridge M, Jackson A. National COVID numbers – Benford's law looks for errors. *Nature* 2020 [access: 19/12/2020];581(7809):384. Available from: <http://www.nature.com/articles/d41586-020-01565-5>
4. Lee K-B, Han S, Jeong Y. COVID-19, flattening the curve, and Benford's law. *Phys A Stat Mech its Appl*. 2020 [access: 19/12/2020];559:125090. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0378437120305719>
5. Sambridge M, Tkalčić H, Jackson A. Benford's law in the natural sciences. *Geophys Res Lett*. 2010;37(22):1-5. DOI: <http://doi.wiley.com/10.1029/2010GL044830>
6. Gómez-Camponovo M, Moreno J, Idrovo AJ, Páez M, Achkar M. Monitoring the Paraguayan epidemiological dengue surveillance system (2009-2011) using Benford's law. *Biomédica*. 2016 [access: 19/12/2020];36(4):583. Available from: <http://www.revistabiomedica.org/index.php/biomedica/article/view/2731>
7. Cabeza García PM. Aplicación de la ley de Benford en la detección de fraudes. *Rev Univ y Soc*. 2019 [acceso: 19/12/2020];11(5):421-7. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2218-36202019000500421
8. Manrique-Hernández EF, Moreno-Montoya J, Hurtado-Ortiz A, Prieto-Alvarado FE, Idrovo AJ. Desempeño del sistema de vigilancia colombiano durante la pandemia de COVID-19: evaluación rápida de los primeros 50 días. *Biomédica*.

2020 [acceso: 19/12/2020];40(Supl. 2):96-103. Disponible en: <https://revistabiomedica.org/index.php/biomedica/article/view/5582>

9. Hurtado-Ortiz A, Moreno-Montoya J, Prieto-Alvarado FE, Idrovo AJ. Evaluación comparativa de la vigilancia en salud pública de COVID-19 en Colombia: primer semestre. *Biomédica*. 2020 [acceso: 19/12/2020];40(Supl. 2):131-8. Disponible en: <https://revistabiomedica.org/index.php/biomedica/article/view/5812>

10. Cinelli C. Benford Analysis for Data Validation and Forensic Analytics. CRAN 2018 [access: 19/12/2020];19. Available from: <https://cran.r-project.org/web/packages/benford.analysis/benford.analysis.pdf>

11. Instituto Nacional de Salud (INS). COVID-19 en Colombia. 2020 [acceso: 9/11/2020];1. Disponible en: <https://www.ins.gov.co/Noticias/Paginas/coronavirus-departamento.aspx>

12. Nigrini M. Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2012. [access: 19/12/2020]. Available from: <http://library.wbi.ac.id/repository/34.pdf>

13. Universidad del Magdalena. Instituto Nacional de Salud autorizó Laboratorio de Biología Molecular de UNIMAGDALENA. 2020 [acceso: 19/12/2020];1. Disponible en: <https://www.unimagdalena.edu.co/presentacionPublicacion/VerNoticia/88061>

Conflicto de intereses

Los autores declaran no tener conflicto de intereses.

Contribuciones de los autores

Jorge Homero Wilches Visbal: Conceptualización, curación de datos, investigación, redacción y revisión.

Midian Clara Castillo Pedraza: Curación de datos, investigación y redacción.

Oskarly Pérez Anaya: Curación de datos, investigación y redacción.