

Datos de precipitación en Ecuador y la Ley de Newcomb-Benford

Paul Jara Ortega¹

¹ Facultad de Ciencias Biológicas, Universidad Central del Ecuador

Corresponding author: Paul Jara Ortega, jarapaul18@gmail.com

(Recibido: 05-08-2023. Publicado: 24-09-2023.)

DOI: 10.59427/rcli/2023/v23.97-101

Resumen

La Ley de Newcomb-Benford (LNB) principalmente se aplica para la detección de datos falsos o erróneos. Existen estudios donde demuestran que datos de precipitación se ajustan a esta ley. En el presente trabajo se trata de conocer si los datos de precipitación mensual en Ecuador se ajustan o no a la distribución de la LNB. Se utilizó 26025 datos de precipitación mensual, provenientes de 411 estaciones meteorológicas del INAMHI, de Ecuador. Con un $\alpha = 0,05$, se aplicó Chi-Cuadrado de Bondad de ajuste, obteniendo un $p = 1,000$. Como el p -valor es mayor al valor α , se aceptó la H_0 : Los datos de precipitación se ajustan a la distribución de la LNB.

Palabras claves: Precipitación, Ley de Newcomb-Benford, Ecuador.

Abstract

The Newcomb-Benford Law (NBL) is mainly applied for the detection of false or erroneous data. There are studies that show that precipitation data conform to this law. In the present work we try to know if the monthly precipitation data in Ecuador fit or not to the distribution of the NBNL. We used 26025 monthly precipitation data from 411 INAMHI weather stations in Ecuador. With an $\alpha = 0.05$, Chi-Square Goodness of Fit was applied, obtaining a $p = 1.000$. Since the p -value is greater than the α value, the H_0 was accepted: Precipitation data fit the LNB distribution.

Keywords: Precipitation, Newcomb-Benford Law, Ecuador.

1 Introducción

La precipitación se define como cualquier tipo de agua meteórica que se encuentra en la superficie terrestre, esto incluye lluvia, granizo, nieve, rocío o escarcha (Sánchez, 2022). El monitoreo de precipitaciones y demás fenómenos meteorológicos e hidrológicos esta encargado el Instituto Nacional de Meteorología e Hidrología (INAMHI) que proporciona la información del tiempo, clima y recursos hídricos del pasado, presente y futuro a la comunidad, con la finalidad de proteger la integridad humana y los bienes materiales (INAMHI, 2023). La LNB se representa como la observación empírica de datos discretos o continuos, determinando que la mayor probabilidad en un grupo de datos es que empiecen con el número 1 y que la menor probabilidad es que empiecen con el 9 (Newcomb, 1881; Benford, 1938), ver figura 1. Esta ley es logarítmica, para el primer dígito significativo (PDS); la expresión matemática de la LNB (Benford, 1938), es:

$$(1) P_{d1} = \log_{10} (1 + d_1^{-1}), \text{ donde } d = 1, 2, \dots, 9$$

(P = Probabilidad del primer dígito significativo).

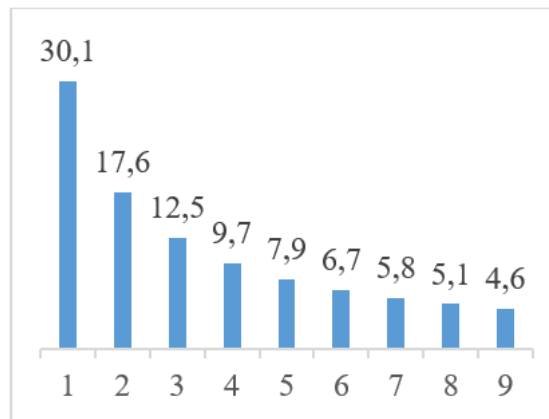


Figure 1: Probabilidades de la LNB para el PDS. Fuente: Newcomb (1881), formato modificado por el autor.

La LNB puede ser aplicada en datos provenientes de una mezcla estadística, en datos no preestablecidos o de origen natural, asimismo los datos deben ajustarse a la LNB sin importar la escala o sistema de medida (Hill, 1995; Gauvrit-Gauvrit y Delahaye, 2008; Nigrini, 2012; Vega-Flores, 2012; Gauvrit-Gauvrit, Houillon y Delahaye, 2017; Cabeza-García, 2019). Para aplicar la LNB se recomienda tener al menos 10000 datos (Morales y Zúñiga, 2010). El uso principal de la LNB es para la detección de fraudes (Cabeza-García, 2019), sin embargo, también existen otras aplicaciones donde los datos se ajustan a esta ley: en datos meteorológicos (Qin et al, 2019; Arias y Poleo, 2022), informática-matemática (Martínez y Canisales, 2009), biología (Costas et al., 2008; Docampo et al., 2009; Campos, Salvo, Flores-Moya, 2016; Cerri, 2018), física (Pain, 2008), datos científicos (Diekmann, 2007) y en esquemas musicales (Khosravani y Rasinariu, 2018). A partir de lo anteriormente descrito el presente trabajo tiene como propósito conocer si los datos de precipitación mensual en Ecuador se ajustan o no a la distribución de la LNB.

2 Metodología

El presente trabajo es descriptivo-analítico. Para ello se utilizó 26025 datos abiertos (Datos abiertos Ecuador: <https://www.datosabiertos.gob.ec/dataset/precipitacion-total-mensual>) de precipitación mensual (PM) en unidades de mm, entre el año 2010 al 2019; procedentes de 411 estaciones meteorológicas del INAMHI (Granda, 2021), ver figura 2.

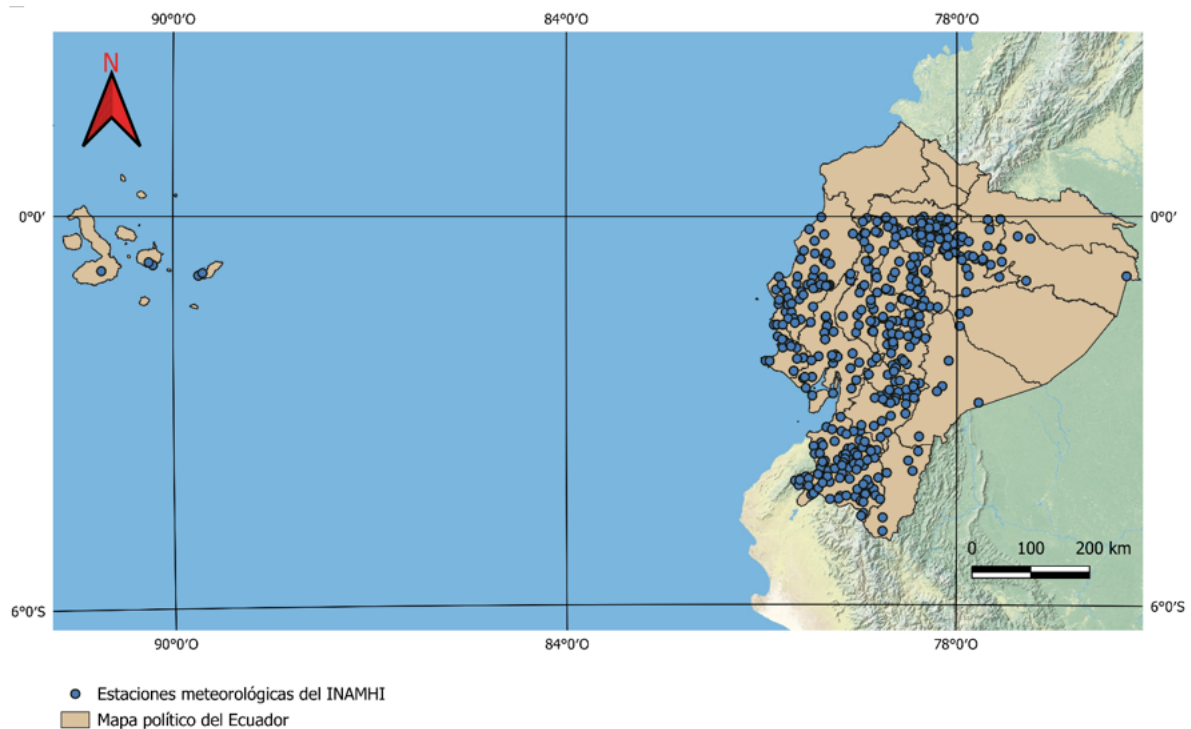


Figure 2: Mapa político del Ecuador con las 411 estaciones meteorológicas. Autor del mapa: Paul Jara-Ortega.(Granda, 2021).

Se realizó una matriz en EXCEL con los 26025 datos, luego se calculó las frecuencias (%) para cada número significativo (1 al 9). Se omitió todos los 0 y celdas sin datos. Para conocer si los datos de precipitación se ajustan a la distribución de la LNB, se aplicó la prueba de Chi-Cuadrado de Bondad de ajuste, utilizando el software Minitab 21. Para ello se planteó la hipótesis nula (p -valor > 0.05): H_0 : Los datos de precipitación se ajustan a la distribución de la LNB.

3 Resultados

Una vez calculadas las frecuencias y porcentajes para el PDS se compararon con los datos de la LNB, ver tabla 1 y figura 3.

Table 1: Distribución de frecuencias para el PDS de las PM, comparadas con los datos de la LNB.

PDS	f	PM%	LNB %
1	7794	29.9481268	30,1
2	4544	17.46013449	17,6
3	3370	12.94908742	12,5
4	2616	10.0518732	9,7
5	2123	8.157540826	7,9
6	1658	6.37079731	6,7
7	1467	5.636887608	5,8
8	1296	4.979827089	5,1
9	1157	4.445725264	4,6

f: frecuencia.

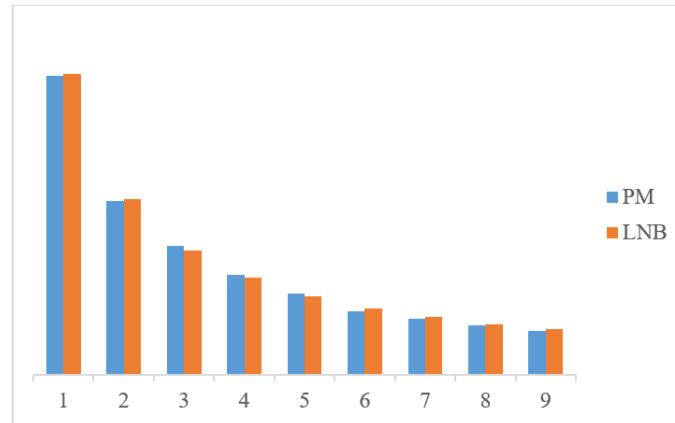


Figure 3: Porcentajes de las frecuencias de las PM comparadas con la LNB.

Con la prueba de Chi-Cuadrado de Bondad de ajuste se determinó los conteos de los datos observados y esperados (tabla 2). Con un $\alpha = 0,05$ se obtuvo un valor Chi-Cuadrado = 0,0679406 y un p -valor = 1,000.

Table 2: Conteos observados y esperados.

Categoría	Observado	Proporción de prueba	Esperado	Contribución a chi-cuadrada
1	29,9481	0,301	30,1	0,0007663
2	17,4601	0,176	17,6	0,0011115
3	12,9491	0,125	12,5	0,0161344
4	10,0519	0,097	9,7	0,0127644
5	8,1575	0,079	7,9	0,0083959
6	6,3708	0,067	6,7	0,0161753
7	5,6369	0,058	5,8	0,0045872
8	4,9798	0,051	5,1	0,0028317
9	4,4457	0,046	4,6	0,0051741

1 (11,11%) de los conteos esperados son menores que 5.

4 Discusión

Como el valor $p = 1,000$ es mayor al valor $\alpha = 0,05$, se acepta la hipótesis nula, indicando que los datos de precipitación se ajustan al patrón logarítmico de la LNB. Esto muestra que el presente trabajo coincide con los trabajos ya realizados en el campo meteorológico (Qin et al, 2019; Arias y Poleo, 2022).

La LNB estudia principalmente posibles fraudes (Cabeza-García, 2019), por lo tanto, si se estudia un conjunto de datos, y éstos se ajustan a la LNB, podemos discernir que los datos son verdaderos, es decir, sin errores o falsificaciones. En caso de datos meteorológicos (Qin et al, 2019; Arias y Poleo, 2022), han podido concluir que sus datos son fidedignos y además mencionar que los equipos de medición se encuentran en buen estado y están arrojando datos fiables. Asimismo, en el presente estudio se pudo comprender que los datos de precipitación mensual son verdaderos, además de describir que, probablemente las estaciones meteorológicas del INAMHI están en buen estado y están proporcionando datos fiables. Se conoce que se necesita al menos 10000 datos para aplicar la LNB, dado que, a mayor cantidad de datos analizados, mayor será el ajuste a la LNB (Morales y Zúñiga, 2010). En el presente estudio se utilizó 26025 datos, lo cual corroboró que, al analizar una cantidad superior a los 10000 datos, estos se ajustaron casi de manera exacta con la distribución de la LNB.

5 Conclusión

Se concluye que los datos de precipitación mensual en Ecuador del 2010 hasta el 2019 si se ajustan a la distribución de la LNB, por ende, la LNB puede aplicarse en estos datos de manera complementaria para confirmar su veracidad, posible falsificación o errores en la toma de datos. También se puede complementar sobre la evaluación del estado de los equipos de medición.

6 Referencias bibliográficas

- Arias Y, Poleo D (2022): Evaluación de la precipitación diaria simulada por el WRF durante 2019-2021 utilizando la ley de Benford. *Tópicos Meteorológicos y Oceanográficos.*, 21:3-24.
- Benford F (1938): The law of anomalous numbers. *Proc. Amer. Philos. Soc.*, 78:551-572.
- Cabeza-García PM (2019): Aplicación de la ley de Benford en la detección de fraudes. *Revista Universidad y Sociedad.*, 11:421-427.
- Campos LA, Salvo E, Flores-Moya A (2016): Natural taxonomic categories of angiosperms obey Benford's law, but artificial ones do not. *Syst. Biodivers.*, 14:431-440.
- Cerri J (2018): A fish rots from the head down: how to use the leading 2 digits of ecological data to detect their falsification. *bioRxiv.*, 3689851:1-14.
- Costas E, López-Rodas V, Toro F J, Flores-Moya A (2008): The number of cells in colonies of the cyanobacterium *Microcystis aeruginosa* satisfies Benford's law. *Aquat. Bot.*, 89:341-343.
- Diekmann A (2007): Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data. *J.Appl.Stat.*, 34:321-329.
- Docampo S, Trigo M, Aira MJ, Cabezudo B, Flores-Moya A (2009): Benford's law applied to aerobiological data and its potential as a quality control tool. *Aerobiologia.*, 25: 275-283.
- Gauvrit-Gauvrit N, Delahaye JP (2008): Pourquoi la loi de Benford n'est pas mystérieuse (Why Benford's law is not mysterious). *Math. & Sci. hum.*, 46:7-15.
- Gauvrit-Gauvrit N, Houillon JC, Delahaye JP (2017): Generalized Benford's Law as a Lie Detector. *Adv. Cogn. Psychol.*, 13:121-127.
- Granda J (2021): Precipitación Total Mensual. Datos abiertos- INAMHI.
- Hill TP (1995): A Statistical Derivation of the Significant-Digit Law. *Statist. Sci.*, 10:354-363.
- Instituto Nacional de Meteorología e Hidrología (INAMHI) (2 de Agosto de 2023): Dirección ejecutiva. Instituto Nacional de Meteorología e Hidrología.
- Khosravani A, Rasinariu C (2018): Emergence of Benford's law in music. *Cornell University. Physics. Soc-ph.*, 2:1-10.
- Martínez R, Canisales C (2009): Ley de Benford y sus aplicaciones. [tesis] Universidad de El Salvador Facultad de Ciencias Naturales y Matemática Escuela de Matemática. Ciudad Universitaria, San Salvador.
- Morales L, Zuñiga M (2010): Sistema utilizando la Ley de Benford para detectar posibles fraudes electorales en las elecciones convocadas en Ecuador. *Escuela Politécnica Nacional. Quito-Ecuador.*
- Newcomb S (1881): Note on the frequency of use of the different digits in natural numbers. *Am. J. Math.*, 4:34-40.
- Nigrini M (2012): Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detections. 1st ed. New Jersey: John Wiley & Sons.
- Pain JC (2008): Benford's law and complex atomic spectra. *Phys. Rev. E.*, 77:1-8.
- Qin L, Han S, Xing L, Zhang Jia, Zhang Jin, Yang R, Wu J (2019): Application Research of Benford's Law in Testing Agrometeorological Data. *IOP Conf. Ser.: Earth Environ. Sci.*, 310:1-11.
- Sánchez F (2022): Hidrología superficial y subterránea. 2ª ed. Kindle Direct Publishing. Salamanca- España.
- Vega-Flores C (2012): "La Ley de Benford y su Aplicación en la Detección de Fraudes Financieros". *Revista Varianza.*, 9:5-7.