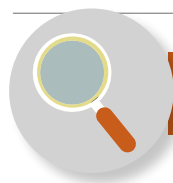


Carlos Gómez ◀ La Ley de Benford y el análisis de información financiera



Perspectiva

La Ley de Benford y el análisis de información financiera

Carlos Gabriel Gómez Villagrán¹

Docente Escuela de Postgrado

Facultad de Ciencias Económicas / USAC

Resumen

La naturaleza de los números puede mostrar que, ante la dicotomía de la vida, puede existir un universo ordenado que, convertido en datos, se puede evaluar si se ajusta a la Ley de Benford. Al tomar una colección de números cualquiera, esta ley indica que los primeros dígitos del uno al nueve no están distribuidos uniformemente, sino que tendrán un sesgo significativo hacia los dígitos más pequeños. La información financiera como tal tiene un comportamiento matemático de una progresión geométrica y dada esta característica puede ser confrontada mediante la Ley de Benford. Por estas razones y ante la cantidad de información financiera que es generada, es necesario tener herramientas de revisión y este patrón cumple como una ellas, de fácil aplicación, verifica los conjuntos de datos observados con estadísticos de prueba y la manipulación de datos se puede realizar con softwares accesibles. La detección oportuna de un error o dato generado intencionalmente, proporciona una ventaja para la toma de decisiones y la gestión de riesgo en cualquier organización o empresa; es por ello que la Ley de Benford se ha convertido en una herramienta para el análisis de datos financieros, que al no cumplirse su patrón propone técnicas más exhaustivas de auditoría.

Palabras clave

Ley de Benford, primer dígito, desviación media absoluta, intervalo de confianza, información financiera.

1. Ingeniero Electrónico (USAC), Maestro en Administración Financiera (USAC), Estudiante de Doctorado en Ciencias Económicas (USAC).

Abstract

The nature of numbers can show that, faced with the dichotomy of life, there can be an ordered universe that, converted into data, can be evaluated if it conforms to Benford's Law. When taking any collection of numbers, this law indicates that the first digits from one to nine are not evenly distributed, but will have a significant bias towards the smallest digits. Financial information as such has a mathematical behavior of a geometric progression and given this characteristic it can be compared by means of Benford's Law. For these reasons and given the amount of financial information that is generated, it is necessary to have review tools and this pattern fulfills as one of them, easy to apply, verifies the observed data sets with test statistics and data manipulation can be performed with accessible software. The timely detection of an error or data generated intentionally, provides an advantage for decision-making and risk management in any organization or company, which is why Benford's Law has become a tool for the analysis of financial data, which, since its pattern is not met, proposes more exhaustive auditing techniques.

Keywords

Benford's Law, First Digit, Mean Absolute Deviation, Confidence Interval, Financial Information.

1. Introducción

¿Será que una solución al cuento de Borges "El jardín de senderos que se bifurcan" (Borges, 2006) donde presenta de una forma metafísica el principio de incertidumbre postulado importante de la mecánica cuántica, es la Ley de Benford? O es más, ante las circunstancias creadas en el mismo cuento, donde los sucesos se bifurcan y los actores ante diversas alternativas escogen unas y eliminan otras, pero esta suposición de escenarios que se bifurcan uno tras otro y la presencia de un Dios ¿no es más que el libre albedrío y la respuesta a este en el universo, no es más que un orden sencillo subyacente contenido en la Ley de Benford?

Pero ¿cómo se puede relacionar esa parte filosófica con la información financiera? Pues nada más que a través de los números; estos, como tales, están en varios ámbitos de la naturaleza y la Ley de Benford proporciona una respuesta sencilla al equilibrio de la aquella. Por ende, siendo la información financiera parte de ese conjunto de cosas naturales, al encontrarse los individuos ante diferentes alternativas en el manejo de esa información, es preciso comparar el comportamiento de los números contenidos en ella, debido a que el orden en la naturaleza funciona de un modo correcto y no hace inconsistencias.

La Ley de Benford no es más que un patrón de números de los primeros dígitos del 1 al 9, que al recolectar un grupo de números generan espontáneamente una estructura la cual dice: que el treinta por ciento de ese grupo de números inicia con el número uno, el diecisiete por ciento con el número dos, y así sucesivamente, hasta indicar que menos del cinco por ciento inicia con el número nueve. Y es aplicable ante cualquier ámbito no solo de la naturaleza, sino del universo.

Por la importancia que ha tenido la evolución de la tecnología, se han abierto nuevas técnicas en la gestión y análisis de datos que dieron lugar a la auditoría analítica forense, que no es más que un camino expedito para lograr estrategias diferentes. Bajo la premisa de que los números no mienten, por medio de la auditoría de datos se revisa la información financiera con el fin detectar errores, tanto voluntarios como involuntarios, invención y alteración de números, maquillaje de datos y fraudes. Esa detección de patrones y tendencias en los números se logra utilizando como herramienta la Ley de Benford, la cual asume que los datos sin error se deben ajustar a ella.

El resto de este estudio se divide en cuatro secciones. En la sección dos se revisa brevemente la Ley de Benford, su origen y base teórica. Así también, se presentan las pruebas de bondad de ajuste. En la sección tres se describe cómo se emplea la Ley de Benford en la información financiera, en el grupo de datos recolectados de muestra. La sección cuatro proporciona los resultados de utilizar la Ley de Benford como herramienta de análisis de datos. Y para finalizar, se presentan las conclusiones.

2. Fundamentos teóricos

Después de referirse varias veces a la Ley de Benford en este estudio, si la intriga del lector lo ha llevado a este apartado, es porque quiere conocer: ¿Qué es?, ¿Cómo funciona?, ¿Quién la descubrió?, ¿Cuál es ese patrón de números? Todos los anteriores cuestionamientos encuentran su respuesta en este apartado, donde se revisa la Ley de Benford desde sus inicios hasta sus aplicaciones, así como su comprobación de ajuste de una forma estadística. Se agregan, en esta sección, las características que cumplen los datos de la información financiera.

2.1. Antecedentes históricos

En 1881 en el observatorio naval de Estados Unidos, el astrónomo Simón Newcomb observó, en los libros de las tablas de logaritmos, que las páginas que contenían los primeros dígitos de dichas tablas eran las más gastadas en comparación con las páginas de los últimos dígitos. Por lo tanto, llegó a la conclusión de que los dígitos que inician con el número 1 eran más frecuentes que los números que comienzan con 2 y así, sucesivamente. Exactamente ¿esto qué

significa? Que los primeros dígitos del 1 al 9 no se distribuyen de una forma uniforme en un conjunto de números, sino que la distribución estará sesgada hacia los primeros números o dígitos más pequeños.

En esa época los logaritmos se utilizaban como una herramienta para el cálculo de operaciones cuando las cantidades eran demasiado grandes; utilizarlos simplificaba enormemente las operaciones debido a que una multiplicación se puede convertir en una suma, una división en una resta, una potencia en una multiplicación; gracias a la ley de los logaritmos, por lo tanto, es más fácil sumar y restar que multiplicar y dividir en grandes números.

Como resultado de su hallazgo, Newcomb escribe el artículo "Note on the Frequency of Use of the Different Digits in Natural Numbers" (Newcomb, 1881), y lo hace publicar en el *American Journal of Mathematics*, pero el texto pasa inadvertido en la sociedad científica de esa época. No es sino en 1938 cuando Frank Benford, que también utilizaba la tabla de logaritmos para realizar sus operaciones de la época en la rama de la física, redescubre lo que anteriormente había observado Newcomb, que era el desgaste de

las primeras páginas de dichas tablas. Para comprobar su hipótesis, Frank llevo a cabo 20 experimentos, recolectando al azar grupos de números de páginas de libros, periódicos, estadísticas de béisbol, longitudes de ríos, constantes matemáticas y físicas, números de casas en las calles, poblaciones, entre otros.

El resultado de sus experimentos fue la realización de una tabla con los datos procesados, y Negrini la vuelve a analizar y presentar (Nigrini, 2012). Con esta evidencia Benford desarrolló las proporciones de los primeros dígitos positivos del 1 al 9, con lo cual obtuvo que la estructura de dichos números fuera una distribución logarítmica. Entonces, Benford publica su artículo “Ley de números anómalos” (Benford, 1938) en *Proceeding of the American Philosophical Society*, el cual tuvo una connotación mayor a la de Newcomb.

Por lo tanto, al comprobarse y postularse por Benford, se tiene que, en una colección o conjunto de datos numéricos recolectados en base de datos como, por ejemplo: tablas, matemáticas, estadísticas y aplicaciones de la vida real, la frecuencia del primer dígito (del 1 al 9) más significativo sigue un patrón particular que no es uniforme.

De tal forma que, de ese grupo de datos, el número 1 es el dígito más significativo y el cual aparece un 30.1% de las veces, le sigue el 2 con una frecuencia de 17.6%; el número 3 con un 12.5%; el 4 a un 9.7%; el 5, con un 7.9%; el 6, un 6.7%; el 7, un 5.8%; el 8, un 5.1% y, finalmente, el 9 con una frecuencia del 4.6%. Esta frecuencia de ocurrencia en los dígitos forma un modelo logarítmico.

Es importante resaltar que al otorgarse el crédito de este descubrimiento a Benford se cumple el axioma propuesto por Stigler en la llamada Ley de la Eponimia de Stigler, la cual indica que: ningún hallazgo científico recibe el nombre de quien lo descubrió en primer lugar (Stigler, 1980).

Hace 30 años aproximadamente, a Marck Nigrini —estudiando su doctorado en la Universidad de Cincinnati— le interesa la llamada Ley de Benford para aplicarla en la detección de fraudes en las declaraciones de impuestos; con esta inquietud da inicio no solo a su tesis doctoral sino a futuras publicaciones, que conducirían a utilizar dicha ley como una herramienta del análisis de datos, sobre todo de la auditoría forense en la información financiera.

2.2. Ley de Benford

Considerando las aseveraciones de Newcomb (Newcomb, 1881) y Benford (Benford, 1938) respecto a que la ocurrencia del primer dígito en cualquier número de un grupo de números, no se distribuye uniformemente, se concluye que:

$$P(D_1 = d_1) = \log_{10} \left(1 + \frac{1}{d_1} \right) \quad (1)$$

Donde: $d_1 = 1 \dots 9$

La ecuación número 1, se conoce como la ley del primer dígito y proporciona la probabilidad (P) de que un número (d_1) tiene de ser el primer dígito significativo de cualquier número, siempre y cuando el primer dígito sea distinto de cero.

Dado que la ecuación 1, proporciona el patrón o mejor dicho una

distribución logarítmica, entonces, se puede calcular la probabilidad de ocurrencia cuando el primer dígito es el número 1; del mismo modo, la probabilidad de que el primer dígito sea el número 2 y así sucesivamente hasta tener el cálculo de que ocurra que el primer dígito sea el número 9. Estos cálculos de las probabilidades de ocurrencia de los primeros dígitos se muestran en la tabla 1.

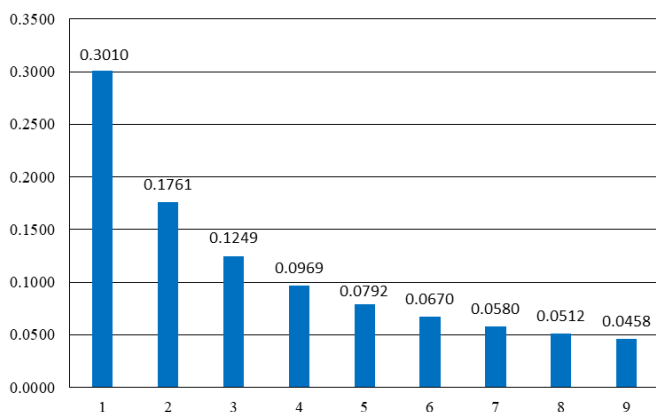
Tabla 1
Proporciones del primero, segundo, tercero y cuarto dígitos de la Ley de Benford

Digito (=d)	Probabilidad del primer digito más significativo	Probabilidad del segundo digito más significativo	Probabilidad del tercer digito más significativo	Probabilidad del cuarto digito más significativo
0	No aplica	0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.10882	1.10097	0.10010
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.09940	0.09994
7	0.05799	0.09035	0.09902	0.0999
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.08500	0.09827	0.09982

Fuente: Benford's Law: Applications for Forensic Accouting, Auditing, and Fraud Detection, M. Nigrini, 2012, p. 6

La figura 1 muestra que los primeros dígitos más significativos no son igualmente probables, lo cual visualiza la afirmación de Newcomb que proporciona una regla exacta en la distribución de los primeros dígitos, que pueden ser observados para muchos conjuntos de números.

Figura 1
Ley de Benford en el primer dígito



Fuente: elaboración propia. El eje Y representa la frecuencia relativa de los primeros dígitos $d1 = 1 \dots 9$. El eje X representa el primer dígito $d1$.

Se hace notar que la probabilidad de ocurrencia del primer dígito en la Ley de Benford, también se puede presentar como proporciones, lo que no es más que una frecuencia relativa de los primeros dígitos. Cabe resaltar que existen varias distribuciones para cada posición de los dígitos en los números (Nigrini, 2012). Pero este estudio se va a centrar en la distribución del primer dígito más significativo, para la muestra recolectada de la información financiera.

Anteriormente se habló de un patrón o estructura que proporciona

la Ley de Benford para los números, la figura 1 muestra ese patrón y una forma sencilla de leer esta gráfica es: en un conjunto de números (datos) recolectados, la probabilidad de ocurrencia de cierto dígito como el primero más significativo disminuye logarítmicamente, a la vez que el valor del dígito aumenta de 1 a 9; por lo tanto, el número 1 aparece como el primer dígito más significativo.

Ahora bien, la ecuación uno se puede generalizar para encontrar la probabilidad de que un dígito en particular se produzca en una

posición determinada en un número. La ecuación presentada abajo muestra cómo se puede calcular la probabilidad de que el dígito 2 se encuentre como segundo dígito y así sucesivamente (Hill, 1995).

$$\sum_{k=10^{n-1}}^{10^n-1} \log_{10} \left(1 + \frac{1}{10k+d_1} \right) = \log_{10} \left[\frac{10k+d_1+1}{10k+d_1} \right] \quad (2)$$

Donde $d_1 = 0 \dots 9$.

Cabe resaltar que la Ley de Benford se puede aplicar a varios ámbitos y áreas de estudio; siempre y cuando, el grupo de datos se recolecte espontáneamente sin ninguna intención; es por ello que no se aplica a los números generados aleatoriamente, pero si al orden generado en la vida cotidiana. Como ejemplos de aplicación puede ser: Longitud de ríos, altura de volcanes, datos de covid, números de calle, notas de música, tamaños en pixeles de fotografías, redes sociales, entre otros.

2.3. Pruebas de comprobación de la distribución de dígitos esperada según la Ley de Benford

Al aplicar la Ley de Benford a un conjunto de datos recolectados, los mismos deben ser verificados si cumplen dicha ley, para lo cual se

procede a comprobar el grado de la desviación de los dígitos de la distribución del grupo de muestra observado, respecto del patrón esperado en la Ley del primer dígito; se utiliza para ello el cortejo de varios estadísticos de prueba.

2.3.1. Chi cuadrada

Básicamente, la Ley de Benford es una distribución de frecuencias y es por esta razón que el estadístico chi cuadrada (χ^2) se adapta bien para someter, a una prueba de hipótesis, las distribuciones de frecuencias.

El estadístico chi-cuadrada es una prueba de bondad de ajuste, que se utiliza para correlacionar variables nominales o cualitativas o, dicho de otra forma, determina la existencia o no de independencia entre dos variables. Esto lo hace calculando la diferencia entre el valor real observado y el valor esperado, y proporciona un número de acuerdo con los grados de li-

bertad de la muestra, el cual se contrasta con el nivel de significancia que sirve como umbral para tomar la decisión de aceptar o rechazar la hipótesis nula.

La distribución chi cuadrada es aquella variable aleatoria continúa X , con n grados de libertad y su función de densidad es dada por:

$$f(x;n)=\begin{cases} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{n/2-1}e^{-x/2}, & 0 < x < \infty \\ 0, & \text{entre otros casos} \end{cases} \quad (3)$$

Donde n es un estero positivo (Tapia Toral, 2018).

La prueba de chi cuadrada, compara la distribución esperada (Ley de Benford) en cada uno de los dígitos, con la distribución observada de la muestra e indica si la distribución observada difiere significativamente de la Ley de Benford, se calcula de la forma siguiente:

$$\chi^2 = \sum_{i=1}^n \frac{(p(o) - p(e))^2}{p(e)} \quad (4)$$

Dónde: $p(o)$ es la frecuencia digito observada, $p(e)$ es la frecuencia del digito esperado acorde a la Ley de Benford, n es el total de dígitos posibles, $i = 1 \dots 9$ (Nigrini, 2012).

5% o 1% que aporta una probabilidad de error; en consecuencia, si el valor p de la chi cuadrado es mayor al nivel de significancia, la hipótesis nula es rechazada.

El cálculo de chi cuadrada proporciona un número que se debe buscar en su tabla a 8 grados de libertad, debido a que solo se está evaluando el primer digito. El resultado se compara con el nivel de significancia que puede ser 10%,

2.3.2. Desviación media absoluta

Es más conocida por sus siglas en ingles MAD, se puede recordar que la desviación es la distancia entre cada dato y su media; bási-

camente, dice qué tan extensos están los valores dentro de un conjunto de datos.

Nigrini indica que la desviación media absoluta (DMA) es menos sensible cuando la muestra es grande, lo que regularmente se tiene para los datos en el mundo real; por lo tanto, esta prueba es ideal porque ignora el número de tamaño de la muestra n . La DMA de un conjunto de datos es la distancia entre cada punto de los datos y la media, y se puede calcular como se muestra a continuación:

$$DMA = \frac{\sum_{i=1}^k |p(o) - p(e)|}{k} \quad (5)$$

Dónde: $p(o)$ es la frecuencia digito observada, $p(e)$ es la frecuencia del digito esperado acorde a la Ley de Benford, n es el total de dígitos posibles, $i = 1 \dots 9$ (Nigrini, 2012).

Con base en ensayos experimentales al tomar varios conjuntos de datos, Nigrini propone unos valores de conformidad para la prueba de desviación media absoluta (DMA) al aplicarla a la Ley de Benford, que se muestran en la tabla 2.

Tabla 2

Valores de desviación media absoluta (DMA) calculados

Rango DMA primero y segundo digito	Conclusión
0.000 a 0.006	Cumple
0.006 a 0.012	Aceptable
0.012 a 0.015	Aceptable marginal
> 0.015	No conformidad

Fuente: Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection, M. Nigrini, 2020.

2.3.3. Z estadística

Para confrontar la distribución real de un conjunto de datos con la distribución esperada, que no es más que la estructura de Benford, también se puede utilizar el estadístico Z en un principio para que funcione de una manera pivote, con el fin de lograr un intervalo de confianza para cada dígito y así poder conocer la desviación que existe entre ambos conjuntos (el real y el esperado). El estadístico Z, es un conjunto de valores que representan cuánto se han dispersado los resultados de la distribución real o de la muestra, por arriba o por debajo de la media de la distribución esperada. Para el caso de la Ley de Benford, el estadístico Z se define de la siguiente forma:

$$Z = \frac{|p(o) - p(e)| - 1/2N}{\sqrt{\frac{p(e)(1 - p(e))}{N}}} \quad (6)$$

Dónde: $p(o)$ es la frecuencia dígito observada, $p(e)$ es la frecuencia del dígito esperado acorde a la Ley de Benford, N es el total de datos de la muestra (Nigrini, 2012).

Para este caso en particular, lo que se requiere es determinar un intervalo de confianza para la frecuencia de cada dígito y se utilizan los valores z como pivotes para obtenerlo. Esto con el fin de dar un intervalo entre un rango (límite inferior y superior) para cada dígito y así poder cortejar la desviación de la muestra observada de la distribución esperada.

Si se recolecta un conjunto de datos como muestra de tamaño N , interesa conocer cierta propiedad de los datos, para este caso la probabilidad de ocurrencia del primer dígito $p = P(D1 = d1)$. Dada la característica anterior, el parámetro p es la proporción de los datos del total de la muestra que verifican la propiedad $D1 = d1$. Sin más que decir, para este caso en particular de la Ley de Benford la variable aleatoria e_i mide a los datos de la muestra (N) la ocurrencia o no de la propiedad $D1 = d1$. Por tanto, e_i sólo toma dos valores: 1 cuando confirma $D1 = d1$

y la probabilidad es p , y 0 cuando no comprueba $D1 = d1$, con probabilidad $1-p$.

Tomando en consideración las características anteriores, e_i es una variable aleatoria cuya distribución es binomial; por lo tanto, tiene parámetros N y p : $X \sim B(n, p)$. Entonces, el intervalo de confianza a construir es para una proporción, siempre considerando como apoyo a Z estadística.

Se puede plantear la ecuación siguiente:

$$P(-z \leq Z \leq z) = 1-\alpha \quad (7)$$

Para el cálculo de los límites superior e inferior y sustituyendo la ecuación 6 en la ecuación 7, y realizando operaciones algebraicas, se tiene lo siguiente:

$$P \left[p(e) - z \sqrt{\frac{p(e) \cdot (1-p(e))}{N}} + \frac{1}{2N} \leq Z \leq p(e) + z \sqrt{\frac{p(e) \cdot (1-p(e))}{N}} + \frac{1}{2N} \right] \quad (8)$$

La ecuación número 8 dice que se han construido 2 estadísticos, que representan el intervalo de confianza [límite inferior, límite superior], en consecuencia, se pueden calcular de la siguiente forma:

Para el límite inferior:

$$p(e) - z \sqrt{\frac{p(e) \cdot (1-p(e))}{N}} + \frac{1}{2N} \quad (9)$$

y para el límite superior:

$$p(e) + z \sqrt{\frac{p(e) \cdot (1-p(e))}{N}} + \frac{1}{2N} \quad (10)$$

Una desventaja del intervalo de confianza con el apoyo de Z , es que a medida que el conjunto de datos es mayor, el estadístico Z tiene desviaciones muy pequeñas; por lo tanto, los límites de intervalo de confianza se colocan muy cerca de la línea de la Ley de Benford (Nigrini, 2011)

2.4. Características de la información financiera

Uno de los entes importantes de este estudio es la información financiera, la cual nace espontánea, se crea debido al funcionamiento de personas, empresas e instituciones públicas, que al tomar un rol de negocios transfieren dinero entre ellos, creando así un valor para su visión y misión. Todos los actores que intervienen de algún modo en la dinámica diaria de los negocios, tienen en común la gestión de la información financiera.

La información financiera que se crea mediante procesos económicos y contables tiene peculiaridades únicas en sus datos, que permiten analizarla de diferentes formas; una de ellas es la forma matemática, pero en este caso es de vital importancia que pueda ser correlacionada con la Ley de Ben-

ford. A continuación, se presenta sus características:

2.4.1. La información contable que se crea, muestra aproximación a una progresión geométrica $S_n = ar^{n-1}$ (Nigrini, 2020) por consiguiente, esta propiedad matemática es muy importante por permitir una relación directa con la Ley de Benford.

2.4.2. La información financiera de muestra puede generarse de una mezcla de distribuciones. Por ejemplo: ventas, gastos de mantenimiento, nóminas, papelería, seguros, entre otros.

Las características anteriores se deben tomar en cuenta en el caso de relacionarlas con la Ley de Benford, dado que al cortejar la información, las cifras o datos sin error deben ajustarse a dicha ley.

3. Metodología

Para el análisis de la información financiera por medio de la Ley de Benford se recolectaron 83,492 datos del rubro de ventas de una empresa, que por motivos de con-

fidencialidad no se puede indicar el nombre (cabe aclarar que en Guatemala es bastante difícil tener acceso a la información), pero para efectos prácticos se llamara "XXXX". Los datos corresponden al período 2019.

La metodología a usar será la estadística descriptiva y seguidamente, para la comprobación de los datos, se utilizarán las pruebas de bondad de ajuste: desviación media absoluta y el intervalo de confianza con la Z estadística.

Se presenta y examina la muestra que es un conjunto de 83,492 números (sin ningún tipo de orden), seguidamente se crea una tabla 3 de análisis de frecuencias, la cual tiene como fin contar la ocurrencia del primer dígito de cada número de la muestra; como resultado se presentan las primeras tres columnas de la tabla llamadas: Primer dígito: números del 1 al 9, frecuencia absoluta (Frec. Abs.) y frecuencia relativa (Frec. Rel.), que no es más que la división entre frecuencia absoluta de cada dígito y la suma total de las frecuencias absolutas.

Se agrega a la tabla 3 la proporción de la Ley de Benford, que se calcula por medio de la ecuación 1. Con todos estos datos se gene-

ra la figura 2, que acota en el eje Y la proporción o frecuencia relativa de los datos observados en este caso, ventas de la empresa "XXXX" y conjuntamente con estos la proporción de la Ley de Benford. En el eje X presenta los primeros dígitos del 1 a 9.

Para comprobar que los datos de ventas de 2019 de la empresa "XXXX" se ajustan a la Ley de Benford, se procede al cálculo de la desviación media absoluta y se corteja el resultado con los valores proporcionados en la tabla 2 y así poder obtener una conclusión. Finalmente, se calcula el intervalo de confianza, se verifica el comportamiento bajo este procedimiento de las desviaciones en cada dígito de la muestra y la Ley de Benford, para presentar finalmente conclusiones.

4. Resultados y discusión

Esta sección es muy importante porque se tiene a la vista la información financiera del rubro de ventas del año 2019 de la empresa "XXXX" y se realizan todos los pasos correspondientes del análisis para cortejar la información y verificar si la misma cumple la Ley de Benford. Así también, se

realiza las pruebas de bondad de ajuste que buscan observar cómo se comportan las desviaciones de los datos del primer dígito respecto a dicha ley.

Como primer paso al análisis con la distribución de la muestra, se genera una *row data*. Cabe hacer notar que de ahora en adelante se le llamara a dicha muestra datos observados.

Como segundo paso, se realiza un análisis descriptivo y éste no es más que el conteo en cada número del primer dígito, como resultado genera la columna de datos

en la tabla 3 llamada Frecuencia absolutas observadas, pero hay que resaltar que el total de datos contados es de 83,488 y en la *row data* se tenía un total de 83,492 números, lo que representa una diferencia de cuatro datos que corresponden a números en el cual su primer dígito es el cero, condición suficiente para rechazarlos.

La tabla 3 muestra los resultados no solo de las frecuencias absolutas observadas, sino que agrega los datos de la frecuencia relativa observada y la proporción de la Ley de Benford.

Tabla 3
Frecuencias relativas distribución observada y Ley de Benford

Primer dígito	Frecuencias absolutas observadas	Frecuencia relativa observada	Proporción Ley de Benford
1	27,114	0.32477	0.30103
2	14,545	0.17422	0.17609
3	10,501	0.12578	0.12494
4	7,687	0.09207	0.09691
5	6,321	0.07571	0.07918
6	5,289	0.06335	0.06695
7	4,561	0.05463	0.05799
8	3,894	0.04664	0.05115
9	3,576	0.04283	0.04576
Total	83,488		

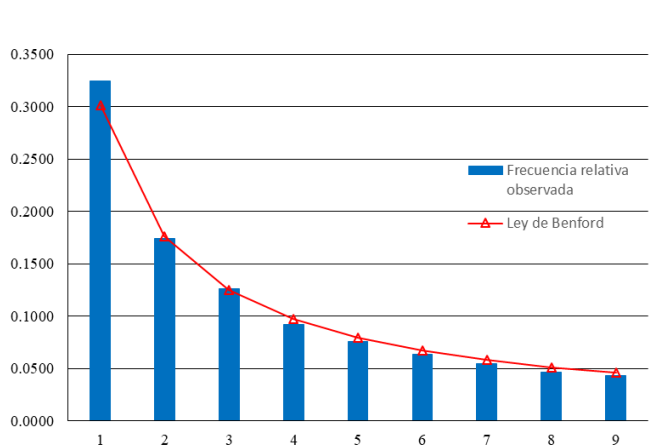
Fuente: Elaboración propia

Este resumen de datos descriptivos es muy útil, porque debido a ellos se genera la figura 2, la cual muestra la frecuencia relativa observada, que no es más que la distribución observada de muestra y a la vez gráfica del patrón de la Ley de Benford.

Al observar esta figura, sin tanto preámbulo, el resultado es que la muestra observada delineada una tendencia en la cual se observa que las barras disminuyen para los dígitos más grandes, de modo que se tiene un patrón cuasi parecido a la ley del primer dígito.

Figura 2

Distribución observada y Ley de Benford del primer dígito



Fuente: elaboración propia. El eje Y representa la frecuencia relativa de los primeros dígitos $d1 = 1 \dots 9$. De la distribución observada y la Ley de Benford. El eje X representa el primer dígito $1 \dots 9$.

Luego de tener a la vista la estadística descriptiva de los datos observados cuantificada en la tabla 3 y mostrados en la figura 2, se dan los resultados de la prueba de bondad de ajuste que, para este

caso con una distribución de datos grande (83,488), se utiliza la desviación media absoluta (DMA) que, según las notas teóricas descritas anteriormente, pueda mostrar mejor ajuste en las desviacio-

nes entre los datos observados y los esperados de acuerdo a la Ley de Benford.

En la tabla 4 se ha anidado una columna que representa la desviación media absoluta (DMA) y en la cual están tabulados todos los resultados operados para cada uno de los primeros dígitos; así también se muestra el resultado de esta prueba el cual es de 0.00546.

Este resultado se contrasta con los intervalos de la tabla 2 (Valores de desviación media absoluta calculados) y se observa que se encuentra dentro de este rango [0.000 a 0.006] donde la conclusión lógica es que cumple o, dicho de otra forma, las desviaciones entre los datos observados y esperados son muy pequeñas y, por lo tanto, no son significativas.

Tabla 4
Desviación media absoluta (DMA)

Primer dígito	Frecuencias absolutas observadas	Frecuencia relativa observada	Proporción Ley de Benford	Desviación media absoluta (DMA)
1	27,114	0.32477	0.30103	0.02374
2	14,545	0.17422	0.17609	0.00187
3	10,501	0.12578	0.12494	0.00084
4	7,687	0.09207	0.09691	0.00484
5	6,321	0.07571	0.07918	0.00347
6	5,289	0.06335	0.06695	0.00360
7	4,561	0.05463	0.05799	0.00336
8	3,894	0.04664	0.05115	0.00451
9	3,576	0.04283	0.04576	0.00292
Total	83,488			0.00546

Fuente: Elaboración propia

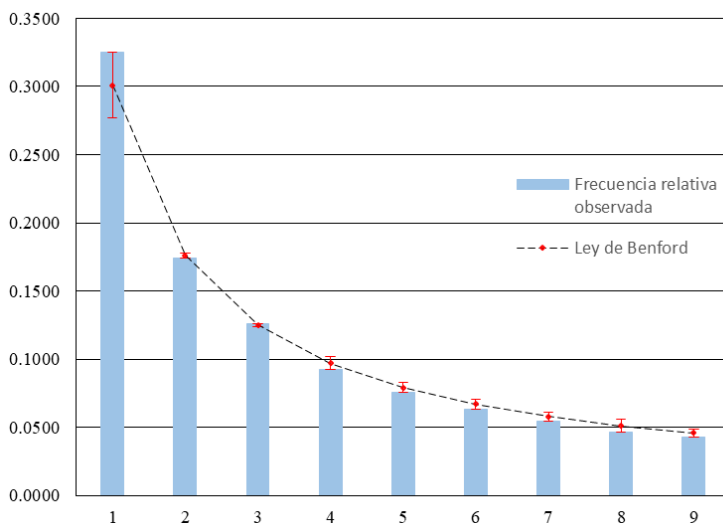
La figura 3 muestra de una forma gráfica las desviaciones para cada primer dígito, con lo cual se pue-

de comprobar que efectivamente las distancias entre los datos observados y la Ley de Benford en la

mayoría de los casos no son significativas; de una forma coloquial, presentan diferencias pequeñas. Pero se debe hacer notar que para el primer dígito más significativo, la desviación entre la muestra observada y la esperada si es significativa. Luego del resultado de esta prueba de bondad de ajuste, se puede indicar que el grupo de muestra observado, que para este caso son las ventas de 2019 de la empresa "XXXX", se ajusta a la Ley de Benford.

Pero cabe resaltar, que siendo esta ley del primer dígito una herramienta de auditoria, que pueda detectar errores, se debe recomendar revisar la información de las cantidades que inician con el número 1. Con esta medida se pueden identificar mejoras en los procesos, así como gestionar mejor el riesgo.

Figura 3
Desviación media absoluta (DMA)



Fuente: elaboración propia. El eje Y representa la frecuencia relativa de los primeros dígitos $d1 = 1 \dots 9$. De la distribución observada y la Ley de Benford. El eje X representa el primer dígito $1 \dots 9$.

A pesar de que no es necesario, dado que ya se cuenta con la prueba de desviación media absoluta, se procede a realizar un intervalo de confianza y como es una muestra muy grande de datos observados se utiliza la Z estadística como se había informado en los fundamentos teóricos. Aunque desde la figura 2 se puede notar que los datos observados cuasi presentan un patrón parecido a la Ley de Benford, se decide realizar un intervalo de confianza y así obtener dos estadísticos para cada primer dígito, así como una Z estadística; los resultados cuantitativos se muestran en la tabla 5, donde para cada primer dígito se tiene un límite inferior y superior que

forman el intervalo de confianza del 95%.

Como la muestra de datos observados es grande, Nigrini (Nigrini, 2012) en sus experimentos con este tipo de muestra indica que a medida que la muestra es más grande, los límites inferior y superior se acercan demasiado a la línea de tendencia de la Ley de Benford. Y como se puede observar en la figura 4 con las líneas rojas punteadas el intervalo de confianza es demasiado pequeño, con lo cual la decisión resulta ambigua, dado que el intervalo no es significativo para muestras muy grandes, como en este caso.

Tabla 5
Intervalo de confianza y Z estadística

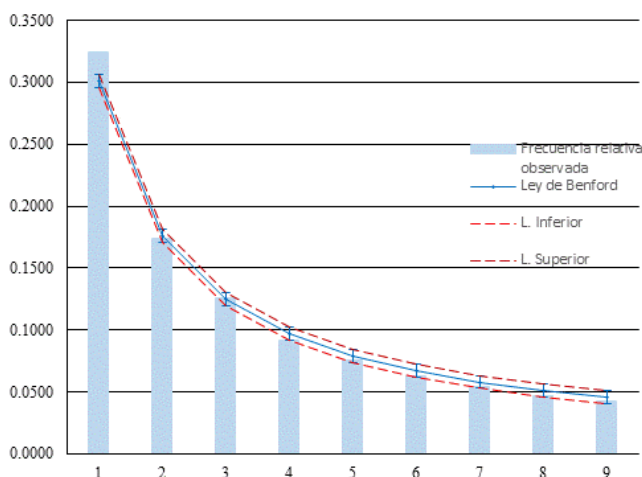
Primer dígito	Frecuencias absolutas observadas	Frecuencia relativa observada	Proporción Ley de Benford	Límite inferior	Límite superior	Z estadística
1	27,114	0.32477	0.30103	0.29531	0.30675	8.13675
2	14,545	0.17422	0.17609	0.17064	0.18154	-0.67439
3	10,501	0.12578	0.12494	0.11964	0.13023	0.31086
4	7,687	0.09207	0.09691	0.09171	0.10211	-1.82334
5	6,321	0.07571	0.07918	0.07405	0.08432	-1.32455
6	5,289	0.06335	0.06695	0.06186	0.07203	-1.38557
7	4,561	0.05463	0.05799	0.05294	0.06304	-1.30414
8	3,894	0.04664	0.05115	0.04613	0.05618	-1.75991
9	3,576	0.04283	0.04576	0.04076	0.05076	-1.14623
Total	83,488					

Fuente: Elaboración propia.

De la tabla 5 se pueden observar los valores de Z estadística; básicamente confirma lo que mostró la prueba de desviación media absoluta y es que el primer dígito más significativo está bastante desviado del patrón mostrado por la Ley de Benford. Para los demás dígitos, los valores de Z estadística muestran pequeñas desviaciones, lo cual confirma que los datos observados son cuasi ajustados a la estructura del Benford.

Como recomendación de la aplicación de esta técnica a la información financiera de ventas de 2019, agregando a lo dicho anteriormente, se puede proceder a revisar a detalle la información para así obtener un mejor modelo de negocio que se base en la aumentar la eficiencia y calidad de los datos.

Figura 4
Intervalo de confianza



Fuente: elaboración propia. El eje Y representa la frecuencia relativa de los primeros dígitos $d1 = 1 \dots 9$. De la distribución observada y la Ley de Benford. El eje X representa el primer dígito $1 \dots 9$.

Al considerar los resultados generales se aprecia que los datos observados cuasi ajustan a la curva de la Ley de Benford; también el valor de la desviación media absoluta está dentro de los parámetros y cumple con ellos como una conclusión lógica. Sobre el intervalo de confianza a un 95% y la Z estadística presentaron resultados no significativos, mostrados tanto cuantitativos como gráficos.

Al no cumplirse la Ley de Benford para el dígito número 1 y dado que los demás números están dentro de las desviaciones establecidas por la prueba de bondad de ajuste desviación media absoluta, la información para el rubro de ventas 2019 de la empresa "XXXX" y en especial toda información donde los montos inicien con 1, deben ser sujetos de revisión.

5. Conclusiones

5.1. La Ley de Benford de una forma general para el universo que es totalmente arbitrario, presenta un orden mediante un modelo matemático simple, que representa un equilibrio para todos los ámbitos.

5.2. El patrón de Benford es una herramienta con meto-

dología simple, que se puede aplicar al análisis de datos de información financiera, para resolver oportunamente y minimizar los riesgos.

5.3. Una información financiera que puede estar corrupta no cumple las tendencias y patrones de la Ley de Benford, lo cual muestra que lo falso no es real.

5.4. Al no cumplir la Ley de Benford, se propone un camino para la revisión minuciosa de los datos o una auditoría eficiente sobre la información financiera.

5.5. Las pruebas de bondad de ajuste para la Ley de Benford, tienen sus limitaciones para grupo de datos muy grandes.

Referencias

Benford, F. (1938) "The law of anomalous numbers". Proc. Amer. Phil. Soc. 78, 551-572.

Borges, J. L. (2006) *Ficciones*. México: Planeta Mexicana.

Douglas A. Lind, S. A. (2012) *Estadística aplicada a los negocios y la economía*. México: McGraw-Hill/ Interamericana Editores

Hill, T. (1995) "A Statistical Derivation of the Significant-Digit Law". *Statistical Science* 10(4), 354-363.

Newcomb, S. (1881) "Note on the Frequency of Use of the Different Digits in Natural Numbers". *American Journal of Mathematics*, 4, 39-40.

Nigrini, M. J. (2011) *Forensic Analytics Methods and Techniques for Forensic Accounting Investigations*. United States of America: Wiley.

Nigrini, M. J. (2012) *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. United States of America: Wiley.

Stigler, S. (1980) "Stigler's law of eponymy". *Transactions of the New York Academy of Sciences* 39, 147-158.

Tapia Toral, M. C. (2018) *Estadística aplicada a la administración y la economía*. Ecuador: CIDE.