“APLICACIÓN DE LA LEY DE BENFORD AL TAMAÑO DE LAS TABLAS DE UNA BASE DE DATOS, Y COMO POSIBLE INDICADOR DE RIESGO INHERENTE DE LA INFORMACIÓN CONTENIDA EN LA MISMA”

AUTORES:

Lic. Héctor Rubén MORALES
(rmorales@eco.unc.edu.ar)

Dra. Cecilia Beatriz DÍAZ
(cdiaz@eco.unc.edu.ar)

Dr. Ricardo Justo CASTELLO
(rcastello@eco.unc.edu.ar)

FACULTAD CIENCIAS ECONOMICAS – UNIVERSIDAD NACIONAL DE CORDOBA

AREA TEMÁTICA: PROYECTOS DE INVESTIGACION

Palabras clave: Benford – registros- tablas - base datos - riesgo

Resumen

La Ley de Benford considera que en un conjunto determinado de números, más del 30% de estos empiezan con el dígito 1, con el dígito 2 inician casi el 18%, y desciende sucesivamente hasta el 9 con sólo el 5%. Este comportamiento ha sido verificado para conjuntos de números que son objeto de estudio en distintos ámbitos científicos. El objetivo de este trabajo es verificar si la distribución estadística de Benford se aplica a los números representados por el tamaño (cantidad de registros) que contienen las distintas tablas que conforman una base de datos.

Los resultados alcanzados confirman esa hipótesis, para lo cual se recurre al análisis estadístico de pruebas de bondad de ajuste.

El estudio pretende servir de base a fin de determinar su uso como posible indicador del riesgo inherente de la información que el auditor utiliza para su control.

1- Introducción

La ley surgida hace casi 80 años del físico Frank Benford se encuentra en pleno estudio y debate sobre su aplicación. Es promovida en distintos ámbitos científicos con la finalidad de verificar si el comportamiento de un conjunto de números, que representan el objeto de estudio, se apega a la misma. Esta ley considera que ciertos dígitos aparecen más frecuentemente que otros en un conjunto determinado de datos. Observa que más del 30% de los números empiezan con el dígito uno, con el dígito dos inician cerca del 18% de los números y, desciende sucesivamente hasta el nueve, a menos del 5% la incidencia como primer dígito de una cifra. También destaca observaciones similares, aunque con frecuencias más estrechas, cuando el análisis se efectúa sobre el segundo dígito del compendio de números estudiados.

Hay distintas posturas que tratan de explicar este patrón de comportamiento que se verifica responde a una función logarítmica (Benford, 1938). Los menos rigurosos sostienen que este fenómeno es sólo la forma en que escribimos los números, como la tesis de Goudsmit y Furry (1944), y hasta la teoría de que ello refleja “la verdad de la naturaleza” (Furlan, 1948).

En los últimos veinticinco años, ciertos estudios analizaron su aplicación sobre datos de la contabilidad con fines de auditoría (Hill, 1995). Consideran que si el comportamiento del conjunto representativo de datos no cumple con la distribución de Benford, puede entenderse la presencia de posibles riesgos de irregularidades o fraudes (Nigrini, 1999).

El objetivo de nuestro trabajo es verificar si la distribución estadística de Benford se aplica al conjunto de números representados por la cantidad de registros (tamaño) que contienen las distintas tablas que conforman una base de datos.

La finalidad primera es la comprobación en sí misma, dado que de la revisión bibliográfica no se advierte un estudio similar. En segundo lugar, se pretende atender –en parte- la problemática o incertidumbre que experimenta el auditor que actúa sobre un sistema de información en un contexto computarizado, respecto a la confianza o posible riesgo inherente en la información contenida en la base de datos que utilizará para su control. Al respecto las Normas Internacionales de Auditoría (NIA) propician que los auditores empleen procedimientos analíticos durante la fase de planificación y ejecución de la auditoría con el objetivo de identificar, entre otras, la existencia de tendencias inusuales.

Para ello se analizan ocho módulos informáticos que contienen datos de entre 3 y 12 años de operatoria de una empresa de envergadura. La metodología aplicada es probar el cumplimiento de la ley de Benford a partir de un análisis puntual (en un momento determinado) sobre la cantidad de registros (tamaño) de las tablas de cada módulo y sobre el conjunto de todos estos módulos informáticos que engloban más de 1.900 tablas y que constituyen la base de datos considerada.

Los resultados alcanzados muestran que la ley de Benford se ajusta a la cantidad de registros de las tablas para el conjunto de todos los módulos que conforman la base de datos. Para ello se recurre al análisis de bondad de ajuste, como la prueba de chi cuadrado y desviación absoluta media.

Entendemos que el presente trabajo resulta novedoso como un posible aporte para contribuir a mitigar la incertidumbre del auditor. En ese sentido, es un estudio empírico que puede servir de base, para un análisis más profundo. El mismo deberá generar el convencimiento acerca de que el resultado alcanzado pueda ser interpretado como un indicador sobre la confianza o alerta del posible riesgo inherente o preexistente de los datos informatizados que son puestos a disposición del auditor al iniciar su tarea de contralor. En consecuencia, este estudio constituye una primera etapa de investigación, pudiendo ser ampliado y/o comparado con otras bases de datos.

2- Revisión bibliográfica

2.1 La Ley de Benford

En 1881, Simon Newcomb, astronómo y matemático, advirtió que su libro de tablas de logaritmos tenía más desgastadas las primeras páginas que las últimas. Dedujo que trabajó más con números cuyas cifras iniciales eran bajas (1, 2 o 3) y menos con aquellas que empiezan con dígitos mayores (7, 8 o 9). Concluyó que la primera cifra de los números no se distribuía de manera uniforme (como podría pensarse), aunque no presentó un argumento formal ni fórmula matemática.

En 1938, y de manera independiente, el físico Frank Benford observó el mismo fenómeno en las tablas de logaritmos y realizó una comprobación empírica sobre un total de 20.229 números agrupados en 20 muestras muy diversas, entre ellas: longitud de más de 300 ríos, cantidad de habitantes de más de 3.200 ciudades, constantes y magnitudes físicas y químicas (como el peso atómico de los elementos), funciones matemáticas e incluso números de direcciones de personas. A partir de los resultados Benford postuló la “ley de los números anómalos” para la probabilidad de que el *primer* *dígito* sea d. Esta ley logarítmica se conoce como “ley de Benford” que se describe como sigue (Benford, 1938):

$$\text{Prob}\,(d_1) = \log_{10}\left(1+\frac{1}{d_1}\right), \quad d_1 = 1, 2, 3,\ldots\ldots,9 \qquad (1)$$

Si la atención es verificar como se distribuye según la ley de Benford el *segundo dígito*, está dada por la siguiente expresión:

$$\text{Prob}(d_2) = \sum_{k=1}^{9}\log_{10}\left(1+\frac{1}{10k+d_2}\right), \quad d_2 = 0,\ 1,\ 2,\ldots\ldots,9 \qquad (2)$$

De igual manera, matemáticamente, se deducen las fórmulas para la ubicación del *tercer dígito* y siguientes.

Los resultados que arroja esta ley respecto a la probabilidad de ocurrencia de los primeros dígitos, se describen en el **Cuadro 1.**

Cuadro 1 – Probabilidad de ocurrencia para cada dígito de acuerdo a la posición que ocupa en un número

Dígito/Posición	Primera	Segunda	Tercera	Cuarta	Quinta o superior
0		11,97%	10,18%	10,02%	10,00%
1	30,10%	11,39%	10,14%	10,01%	10,00%
2	17,61%	10,88%	10,10%	10,01%	10,00%
3	12,49%	10,43%	10,06%	10,01%	10,00%
4	9,69%	10,03%	10,02%	10,00%	10,00%
5	7,92%	9,67%	9,98%	9,99%	10,00%
6	6,69%	9,34%	9,94%	9,99%	10,00%
7	5,80%	9,04%	9,90%	9,99%	10,00%
8	5,12%	8,76%	9,86%	9,99%	10,00%
9	4,58%	8,50%	9,83%	9,98%	10,00%

Fuente: elaboración propia a partir de la Ley de Benford.

Es decir, según esta ley, el 30,1% de las veces, la primera cifra significativa (no incluye el 0) será un 1, mientras (en el otro extremo) sólo un 4,6% de las veces será 9.

Esta ley presenta una propiedad matemática que la hace exclusiva. Es la única ley de probabilidad invariante frente a cambios de escala. Se aplica independientemente de la escala de medición. Arribamos al mismo resultado tanto si trabajamos o convertimos la información multiplicando o dividiendo por una constante, o bien usando datos en kilómetros, millas o metros, o en términos financieros daría lo mismo utilizar importes en Pesos o Dólares.

Mientras como limitantes principales se cuenta que no es aplicable a un conjunto de números aleatorios como la lotería, o de números asignados, como los números de teléfono celular, las cédulas de identidad o números de cheques, ya que comienzan con números correlativos; o números que fluyen sólo en un rango determinado, como la estatura de las personas.

2.2 Antecedentes de aplicaciones de la Ley de Benford

La ley de Benford está siendo profundamente investigada en distintos ámbitos científicos, sobre los cuales a continuación se citan algunos trabajos sólo a título ilustrativo. Desde demostrar que se ajusta a las constantes físicas (Burke y Kincanon, 1991), hasta su aplicabilidad sobre el comportamiento estadístico y posible detección de terremotos (Sambridge, et al. 2011). También se analiza como indicador de fiabilidad para evaluar riesgo de toxicidad (Pepijn de Vries et al., 2013), la posible detección de fraudes electorales (Roukema, 2009 y Castañeda, 2011), y el análisis sobre cantidad de seguidores de distintas redes sociales (Golbeck, 2015).

En general, la mayoría de los desarrollados antes mencionados emanaron luego de que fuera estudiada su aplicación en las ciencias económicas, y encontrara espacio concreto para su utilidad. Varian (1972), economista, sugiere que la ley de Benford puede usarse como una prueba de la honestidad o validez de datos científicos supuestamente aleatorios en un contexto de ciencias sociales. Esto recién fue recogido por los contables a fines de los años ochenta. En ese momento, dos estudios se basaron en el análisis digital para detectar la manipulación de los ingresos. Carslaw (1988) encontró que el número de ganancias de las empresas neozelandesas no se ajustaba a la distribución esperada, y a su vez Thomas (1989) descubrió un patrón similar en las ganancias de las empresas estadounidenses.

Hill (1995), proporcionó una prueba para la ley de Benford, y demostró cómo se aplicaba a los datos bursátiles, las estadísticas del censo y ciertos datos contables. Señaló que la distribución de Benford, como la distribución normal, es un fenómeno observable empíricamente.

Nigrini parece ser el primer investigador en aplicar la ley de Benford de manera amplia a los números de la contabilidad y con el objetivo de detectar posibles fraudes. Para su tesis tomó el trabajo sobre la manipulación de ganancias por Carslaw y Thomas, y añadió el de Benford. Su estudió se basó en el análisis digital para ayudar a identificar a los evasores de impuestos (Nigrini 1996). Posteriormente, publicó distintos artículos que detallan aplicaciones prácticas con fines de auditoría a partir de pruebas sobre conjuntos de números de contabilidad (Nigrini y Mittermaier 1997, Nigrini 1999).

No obstante, la literatura académica es algo cautelosa al hacer afirmaciones sobre la efectividad de los procedimientos basados en la ley de Benford para detectar el fraude. En general, se sostiene que si al someter a prueba un conjunto de datos, este no se ajusta a la ley de Benford, sólo puede mostrar ineficiencias operativas o fallas en sistemas, en lugar de precisar un fraude (Etteridge y Srivastava, 1999). Bajo esa óptica se la entiende, entonces, como una orientación concreta a dónde fijar el énfasis del control.

Por su parte, la normativa internacional insta al auditor que actúa en entornos informatizados a evaluar la confiabilidad de los datos con los que desarrolla su tarea y a aplicar pruebas de funcionamiento de controles y pruebas sustantivas sobre datos a nivel de transacciones (NIA 315, 330, 401). En tal sentido, cierta literatura propicia para ello, entre otros, el análisis de Benford.

Nuestro trabajo trata de complementar esa propuesta, aplicando Benford en una etapa preliminar a aquella del control a nivel de datos puros. Nos referimos a considerar este control previo, como un aporte al denominado (por la normativa, NIA 400) riesgo inherente de la información que es puesta a disposición del auditor para su contralor.

El riesgo inherente está dado en las características de la entidad y del sistema de información a ser analizado. Este riesgo no puede ser cambiado por el auditor pero si conocerlo, es innato de la empresa. Es una medida apriorística del riesgo, la cual es independiente a los controles que se estén aplicando. Es una evaluación por el auditor de las posibilidades de encontrar errores importantes antes de la evaluación del control interno (Informe 16, FACPCE, 2009).

Verificar si el perfil de la base de datos se ajusta a la distribución de Benford, a partir de analizar el tamaño (cantidad de registros) de las tablas que la componen, permitiría esa mirada anticipada o previa, superficial pero a la vez casi radiográfica de la base de datos. Este diagnóstico ayudaría a mitigar la incertidumbre del auditor, como posible indicador de fiabilidad, utilidad esta que se aplica en otras disciplinas.

3. Objetivo del estudio:

El objetivo principal es verificar si la distribución estadística de Benford se aplica al conjunto de números representados por el tamaño (cantidad de registros) que contienen las distintas tablas que conforman una base de datos relacional.

El objetivo secundario es servir de base para profundizar el análisis, a fin de determinar su uso como posible indicador de riesgo inherente de la información que el auditor utiliza para su control.

4. Metodología

4.1 Fuente de datos

A los fines del análisis empírico para verificar si la ley de Benford se ajusta a la cantidad de registros de las tablas de una base datos, se recurrió a los datos de una empresa de envergadura que opera en el sector energético del país, prestando servicio a más de 1,1 millones de clientes. La base de datos es relacional bajo Oracle, y cuenta con distintos módulos funcionales, según la especificidad del proceso (comercial, contabilidad, sueldos, etc).

Los datos obtenidos que constituyen el perfil de la base de datos, se refieren solo a la cantidad de registros que contienen las tablas que componen cada módulo a un momento determinado, el que coincide con la ejecución de un back up. Se excluyeron del análisis las tablas nulas, que no cuentan con registros.

Para conformar la base de datos, se seleccionaron ocho módulos informáticos, tratando de que la composición resulte heterogénea. La heterogeneidad se basó en considerar módulos de características diferentes en cuanto al tipo de procesos que realizan como en el volumen de la información que almacenan. De este modo, entendemos que generamos una mayor complejidad para el análisis y en consecuencia, una mayor robustez en los resultados a que se arriben.

En el **Cuadro 2** se describe la composición de cada uno de los módulos que integran la base de datos Esta contiene un total de 1.923 tablas con casi 4.500 millones de registros. Allí se puede apreciar los distintos volúmenes en tablas y cantidad total de registros por cada módulo. El módulo de Gestión Comercial (GC) es el de mayor trascendencia con 808 tablas (42%) que alcanza al 91% de los datos. En relación inversa se observa que Recursos Humanos (RH) tiene 340 tablas (18%) que contienen sólo el 3% de los datos.

Cuadro 2 - Módulos informáticos que conforman la Base de Datos

Módulo informático	Cantidad tablas	Total registros módulo
Contabilidad (CO)	78	225.245.009
Recursos Humanos (RH)	340	1.483.944
Inventario (IN)	71	14.916.459
Gestión Obras (GO)	58	380.572
Solicitudes Int. (SI)	87	1.832.473
Gestión Comercial (GC)	808	4.093.412.030
Liquidación sueldos (SU)	381	135.126.395
Gestión Expedientes (EX)	100	13.416.811
Total Base de Datos	1.923	4.485.813.693

Fuente: elaboración propia.

4.2 Clasificación de los datos

Mediante Excel se logró aislar el primer dígito con el cual inicia la cantidad de registros que contiene cada tabla. Luego, mediante tablas dinámicas se obtiene el **Cuadro 3**, donde se describe por cada módulo la suma de tablas cuyas cantidades de registros inician con el dígito 1 y hasta el 9. Ejemplo: el módulo Contabilidad (CO) tiene 78 tablas. De esas, hay 22 tablas cuya cantidad de registros inicia con el dígito1, luego hay 11 tablas donde la cantidad de registros inicia con el dígito 2, y así sucesivamente.

Cuadro 3 - Discriminación en función al dígito con que inician las cantidades de registros de las tablas de cada módulo

DIGITO	CO	RH	IN	GO	SI	GC	SU	EX	Total
1	22	94	21	12	23	254	114	26	**566**
2	11	74	17	13	17	135	70	24	**361**
3	11	35	11	12	10	89	62	13	**243**
4	6	38	6	2	10	89	39	12	**202**
5	5	34	7	4	9	54	33	6	**152**
6	10	20	2	6	2	60	23	7	**130**
7	4	16	3	5	5	39	14	7	**93**
8	5	15	3	3	5	45	14	3	**93**
9	4	14	1	1	6	43	12	2	**83**
Total	**78**	**340**	**71**	**58**	**87**	**808**	**381**	**100**	**1.923**

Fuente: elaboración propia.

4.3 Análisis de los datos

En primer término, y previo al inicio del análisis propiamente dicho, corresponde revisar los posibles limitantes que rigen para aplicar Benford a un conjunto de números.

En general se cumple con todos los condicionantes que considera la literatura. El conjunto de datos está formado por magnitudes medibles de un mismo fenómeno. Los datos no son números asignados o aleatorios. La distribución de la variable es ligeramente asimétrica positiva, es decir tiene un mayor número de valores pequeños que grandes, lo que es consecuencia natural del fenómeno analizado (cantidad de registros). Los datos están generados en periodos de tiempo muy prolongados, dado que entre los módulos el rango de la información oscila entre 3 y 12 años.

Por otra parte, la bibliografía recomienda que el tamaño del conjunto de datos sea mayor a 1.000, lo que permite establecer conclusiones de auditoría para la prueba del primer dígito. Sobre este aspecto, se cumple respecto a la cantidad de tablas que conforman la base de datos en su conjunto. Pero si el análisis se efectúa por cada módulo, al reunir una cantidad menor de tablas podría resultar que no cumpla con la distribución de Benford. No obstante, según Nigrini cita que el estudio de Wallace (2002), se verifica el cumplimiento de Benford usando cuatro conjuntos de datos con sólo 67 observaciones cada uno (Nigrini et al., 2007).

En segundo término, en **Cuadro 4** procedemos a convertir para cada módulo las frecuencias del número de tablas (expuestas en el anterior Cuadro) y por cada dígito en valores porcentuales. De esta manera logramos que nos resulten comparables con los porcentajes esperados según la distribución de Benford para el primer dígito.

Cuadro 4 - Detalle porcentual por dígito con que inician las cantidades de registros de las tablas de cada módulo

DIGITO	CO	RH	IN	GO	SI	GC	SU	EX	Total	Benford
1	28,2%	27,6%	29,6%	20,7%	26,4%	31,4%	29,9%	26,0%	**29,4%**	**30,1%**
2	14,1%	21,8%	23,9%	22,4%	19,5%	16,7%	18,4%	24,0%	**18,8%**	**17,6%**
3	14,1%	10,3%	15,5%	20,7%	11,5%	11,0%	16,3%	13,0%	**12,6%**	**12,5%**
4	7,7%	11,2%	8,5%	3,4%	11,5%	11,0%	10,2%	12,0%	**10,5%**	**9,7%**
5	6,4%	10,0%	9,9%	6,9%	10,3%	6,7%	8,7%	6,0%	**7,9%**	**7,9%**
6	12,8%	5,9%	2,8%	10,3%	2,3%	7,4%	6,0%	7,0%	**6,8%**	**6,7%**
7	5,1%	4,7%	4,2%	8,6%	5,7%	4,8%	3,7%	7,0%	**4,8%**	**5,8%**
8	6,4%	4,4%	4,2%	5,2%	5,7%	5,6%	3,7%	3,0%	**4,8%**	**5,1%**
9	5,1%	4,1%	1,4%	1,7%	6,9%	5,3%	3,1%	2,0%	**4,3%**	**4,6%**
Total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Fuente: elaboración propia.

Observando los valores del Cuadro 4 y comparando con la distribución de cada módulo con la de Benford para el primer dígito (última columna), se pueden advertir a simple vista pocas coincidencias y muchos desvíos. Entre los desvíos, algunos aparentan ser significativos. Un ejemplo puede ser en el módulo CO, donde el dígito 6 como primera cifra de la cantidad de registros, está en el 12,8% de las tablas que lo componen, cuando lo esperado según Benford es 6,7%. Sobre esta cuestión, podría tenerse presente que puede ser una consecuencia de lo comentado en párrafos anteriores. Es decir, que puede resultar un limitante la cantidad baja de datos analizados, como presenta el módulo CO que comprende 78 tablas en total.

No obstante, consideramos que en esta primera instancia del análisis debemos ir de lo general (la base de datos) a lo particular (cada módulo).

Así, en tercer término, nos centramos en comparar y analizar los valores obtenidos para el Total de la base de datos con Benford, es decir, las dos últimas columnas del Cuadro 4. Si bien de la simple comparación se observan diferencias estas parecer no ser muy significativas.

4.4 Pruebas de bondad de ajuste

Un análisis más preciso nos provee la prueba de bondad de ajuste Chi Cuadrado. La hipótesis nula (Ho) es que los datos reales u observaciones siguen la distribución de probabilidad esperada por la Ley de Benford. La fórmula de Chi Cuadrado (χ^2) considerada es la siguiente:

$$\chi^2 = \sum_{d=m}^{9} \frac{(P_{obs}(d) - P_t(d))^2}{P_t(d)} \qquad (3)$$

donde:
- *Pt(d)* es la frecuencia esperada según Benford
- *Pobs(d) es la frecuencia observada*
- *m* es el dígito analizado. En este estudio es sólo el primer dígito (*m*=1)

Para su aplicación recurrimos a las frecuencias observadas del total de la base de datos reales (ya conocidas, Cuadro 3), y a las frecuencias esperadas, que surgen de considerar la distribución de Benford para cada dígito sobre el número 1923 que son la totalidad de las tablas. Así, por ejemplo, la frecuencia esperada para el dígito 1, es 578,88 (1923 * 0,301) y para el dígito 2 es 338,62 (1923 * 0,176).

En función a lo expresado, conformamos el **Cuadro 5**, y obtenemos los componentes de χ^2:

Cuadro 5 - Obtención de χ^2 a partir de las frecuencias reales y las esperadas de la base de datos

DIGITO	Observ. *Pobs(d)*	Benford *Pt(d)*	*Pobs(d) - Pt(d)*	*(Pobs(d) - Pt(d))²*	*(Pobs(d) - Pt(d))²* / *Pt(d)*
1	566	578,88	-12,88	165,91	0,29
2	361	338,62	22,38	500,71	1,48
3	243	240,26	2,74	7,52	0,03
4	202	186,36	15,64	244,67	1,31
5	152	152,27	-0,27	0,07	0,00
6	130	128,74	1,26	1,59	0,01
7	93	111,52	-18,52	342,94	3,08
8	93	98,37	-5,37	28,80	0,29
9	83	87,99	-4,99	24,92	0,28
Total	1.923	1.923	**Valor de Chi Cuadrado (χ^2) ->**		**6,77**

Fuente: elaboración propia.

Luego determinamos el valor crítico para la distribución de χ^2 para un valor de α = 0,05 (confianza del 95%) y 8 grados de libertad ((9 filas – 1) x 1 columna). Ese valor es $\chi^2_{(0,95,8)}$ = 15,51 (en Excel como PRUEBA.CHI.INV(0,05;)= 15.5073). Al ser el estadístico obtenido menor al valor crítico de 15,51 se acepta que los datos se ajustan a la ley de Benford.

Luego, nos ocupamos de hacer la prueba de ajuste χ^2 para cada uno de los módulos. De igual manera se conforma el **Cuadro 6**, con las frecuencias reales (obtenidas en el Cuadro 3) y las esperadas según la distribución de Benford. Por ejemplo, para el módulo CO, en el dígito 1 la frecuencia esperada es 23,5 (78 * 0,301).

Cuadro 6 - Frecuencias reales y esperadas según Benford para cada módulo

Dig.	CO		RH		IN		GO		SI		GC		SU		EX	
	Real	Esper	Real	Esper.	Real	Esper	Real	Esper.	Real	Esper	Real	Esper.	Real	Esper	Real	Esper.
1	22	23,5	94	102,3	21	21,4	12	17,5	23	26,2	254	243,2	114	114,7	26	30,1
2	11	13,7	74	59,9	17	12,5	13	10,2	17	15,3	135	142,3	70	67,1	24	17,6
3	11	9,7	35	42,5	11	8,9	12	7,2	10	10,9	89	100,9	62	47,6	13	12,5
4	6	7,6	38	32,9	6	6,9	2	5,6	10	8,4	89	78,3	39	36,9	12	9,7
5	5	6,2	34	26,9	7	5,6	4	4,6	9	6,9	54	64,0	33	30,2	6	7,9
6	10	5,2	20	22,7	2	4,7	6	3,9	2	5,8	60	54,1	23	25,5	7	6,7
7	4	4,5	16	19,7	3	4,1	5	3,4	5	5,0	39	46,9	14	22,1	7	5,8
8	5	4,0	15	17,4	3	3,6	3	3,0	5	4,5	45	41,4	14	19,5	3	5,1
9	4	3,6	14	15,6	1	3,3	1	2,7	6	4,0	43	37,0	12	17,4	2	4,6
Total	78	78	340	340	71	71	58	58	87	87	808	808	381	381	100	100
χ^2	**6,09**		**9,48**		**6,15**		**10,98**		**5,17**		**8,55**		**11,34**		**6,51**	

Fuente: elaboración propia.

Los valores obtenidos de χ^2 son menores al valor crítico de 15,51 por lo que todos los módulos se ajustan a la Ley de Benford.

Cabe aclarar que otra alternativa para ejecutar χ^2 para la base de datos total sería la suma de todos los χ^2 obtenidos. Esto arroja para toda la base de datos un χ^2=64.28. A su vez, el valor crítico χ^2 será para 56 grados de libertad ((9 filas – 1) * (8 columnas – 1)), que es de 74,47. Al ser el valor crítico mayor que el estadístico obtenido, también se confirma que el comportamiento de las frecuencias reales para toda la base de datos se ajusta a la Ley de Benford.

Por otra, se debe apuntar que el desarrollo anterior sobre el cálculo de χ^2 se hizo discriminado en etapas a fin de ilustrar sobre el método. No obstante, otra forma sería a través del estadístico Z y/o recurriendo a software como STATA o SPSS entre otros.

Nigrini (et al. 2012) considera que corresponde también aplicar el análisis de bondad mediante el test de la desviación absoluta media (MAD). La fórmula es:

$$MAD = \frac{1}{9}\sum_{d=1}^{9} |P_{obs}(d) - P_t(d)| \qquad (4)$$

donde: - *Pt(d)* es la proporción esperada según Benford

- *Pobs(d) es proporción observada*

Cuando se utiliza este estadístico para la Ley de Benford, Nigrini (et al. 2012) sostiene que se puede determinar el nivel de conformidad según el rango donde se encuentren los valores obtenidos. Estos se muestran en **Cuadro 7.**

Cuadro 7 – Rangos de Conformidad para MAD

Rango	Nivel de Conformidad
0.000 a 0.006	Alta
0.006 a 0.012	Acepta
0.012 a 0.016	Media
Más de 0.016	Baja

En función a ello, y partiendo de los datos del Cuadro 3, efectuamos los cálculos de la MAD y calificamos siguiendo el criterio de Nigrini. Esto se expone en **Cuadro 8**.

Cuadro 8 - Cálculo del MAD para cada módulo y para la base de datos de las tablas de cada módulo

DIGITO	CO	RH	IN	GO	SI	GC	SU	EX	Total	Benford
1	28,2%	27,6%	29,6%	20,7%	26,4%	31,4%	29,9%	26,0%	**29,4%**	**30,1%**
2	14,1%	21,8%	23,9%	22,4%	19,5%	16,7%	18,4%	24,0%	**18,8%**	**17,6%**
3	14,1%	10,3%	15,5%	20,7%	11,5%	11,0%	16,3%	13,0%	**12,6%**	**12,5%**
4	7,7%	11,2%	8,5%	3,4%	11,5%	11,0%	10,2%	12,0%	**10,5%**	**9,7%**
5	6,4%	10,0%	9,9%	6,9%	10,3%	6,7%	8,7%	6,0%	**7,9%**	**7,9%**
6	12,8%	5,9%	2,8%	10,3%	2,3%	7,4%	6,0%	7,0%	**6,8%**	**6,7%**
7	5,1%	4,7%	4,2%	8,6%	5,7%	4,8%	3,7%	7,0%	**4,8%**	**5,8%**
8	6,4%	4,4%	4,2%	5,2%	5,7%	5,6%	3,7%	3,0%	**4,8%**	**5,1%**
9	5,1%	4,1%	1,4%	1,7%	6,9%	5,3%	3,1%	2,0%	**4,3%**	**4,6%**
Total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
MAD	**0,0213**	**0,0172**	**0,0251**	**0,0434**	**0,0203**	**0,0102**	**0,0130**	**0,0238**	**0,0049**	
Conform	***Baja***	***Baja***	***Baja***	***Baja***	***Baja***	***Acepta***	***Media***	***Baja***	***ALTA***	

Fuente: Elaboración propia.

5. Resultados:

Los resultados obtenidos por el test de la χ² permiten afirmar que, considerando el primer dígito de la distribución de las cantidades de registros que contiene las tablas de la base de datos analizada, siguen la Ley de Benford. También las distribuciones de las observaciones de cada módulo informático permiten confirmar que se asemejan a lo esperado por la Ley de Benford. Es decir, en todos los casos a través de χ² se acepta la hipótesis nula.

El test de la Desviación Absoluta Media (MAD) presenta resultado muy favorable para la base de datos en su conjunto, cumpliendo con la Ley de Benford. Cuando el MAD se realiza a nivel de cada módulo, surge que solo los módulos GC (Gestión Comercial) y SU (Sueldos) permiten afirmar la hipótesis de que la distribución de sus datos se asemejan a la Ley de Benford. En tanto en los seis módulos restantes se rechaza la hipótesis nula.

Profundizando el resultado del MAD, se observa que, además de la base de datos en su conjunto, los módulos GC y SU, que cumplen con la Ley de Benford, son los de mayor cantidad de tablas y, están entre los de mayor volumen de información. Esto se puede apreciar en el **Cuadro 9**.

Cuadro 9 – Módulos informáticos ordenados en función a menor MAD y cantidad de tablas que contienen

Módulo informático	Cantidad tablas	Total registros módulo	MAD Valor	MAD Conformidad
Total BD	1.923	4.485.813.693	**0,0049**	**ALTA**
GC	808	4.093.412.030	**0,0102**	**BUENA**
SU	381	135.126.395	**0,0130**	**MEDIA**
RH	340	1.483.944	**0,0172**	**BAJA**
SI	87	1.832.473	**0,0203**	**BAJA**
CO	78	225.245.009	**0,0213**	**BAJA**
EX	100	13.416.811	**0,0238**	**BAJA**
IN	71	14.916.459	**0,0250**	**BAJA**
GO	58	380.572	**0,0434**	**BAJA**

Fuente: elaboración propia.

En función a lo analizado, se observa en general que a menor cantidad de tablas del módulo, el valor del MAD es mayor, lo que lleva a la no conformidad y rechazar la hipótesis nula. Esta cuestión, como se dijo anteriormente, se puede corresponder con los posibles limitantes para verificar el comportamiento de Benford ante una baja cantidad de datos (número de tablas del módulo).

6. Conclusión:

En primer lugar, como experiencia de este estudio, comprendemos el potencial de análisis que encierra el fenómeno de la distribución de la Ley de Benford. En este sentido, adherimos a la postura de Furlan (1948) al afirmar que refleja “la verdad de la naturaleza”. A ello, desde nuestra postura inicial, le sumamos que esa verdad se observa tanto en la naturaleza estática (longitud de ríos o constantes matemáticas), como en la naturaleza que fluye y crece día a día (poblaciones de ciudades, el número de seguidores de redes sociales o el tamaño de las tablas de una base de datos).

De acuerdo con el análisis realizado y los resultados alcanzados, se verifica que la ley de Benford se ajusta a la distribución del tamaño (en cantidad de registros) de las tablas que integran una base de datos.

Si el análisis se parcializa a nivel de los módulos que conforman la base de datos, el ajuste con la distribución de Benford se verifica con la prueba de bondad de ajuste de Chi Cuadrado. No ocurre lo mismo con la prueba de ajuste mediante la prueba de Desviación Absoluta Media (MAD). En este último, la prueba del MAD muestra que el ajuste con Benford no se cumple cuando la cantidad de tablas del módulo es menor a 350 tablas. El MAD alcanza una conformidad media con casi 400 tablas y valores aceptables con 800. La extrapolación indica que a partir de 700 tablas, es la cantidad mínima para alcanzar un ajuste aceptable

Dado que la cantidad de tablas de un módulo es de baja elasticidad, consideramos que el estudio podría ampliarse tratando de adecuarse para salvar esos limitantes. En tal sentido, la propuesta, ya testeada en parte, es considerar no solo un análisis en un momento determinado, sino sumar a ese punto de partida otros momentos (otros back up). La suma de cada momento duplicaría la cantidad de tablas pero con distintas cantidades de registros por el propio crecimiento. Esto permitiría revisar la distribución de Benford según la evolución temporal de la base de datos.

De lograrse resultados aceptables, resultaría posible indagar en la factibilidad de establecer un orden de prioridad sobre los módulos en que se debe orientar la atención de la auditoría o área de control.

Finalmente, consideramos que el presente trabajo es novedoso y puede ser un aporte para contribuir a mitigar la incertidumbre del auditor. Lo expresado se basa en la experiencia alcanzada en otras disciplinas, que también han demostrado que el conjunto de números analizados cumplen con la distribución de Benford. Estas infieren y algunas lo demuestran, que si la distribución sometida a examen no se ajusta a Benford, existen indicios de posibles irregularidades. Consideramos que este tipo de estudio empírico puede servir de base, para un análisis más profundo. El mismo deberá generar el convencimiento acerca de que el resultado alcanzado pueda ser interpretado como un indicador sobre la confianza o alerta del posible riesgo inherente o preexistente de los datos informatizados que son puestos a disposición del auditor al iniciar su tarea de contralor. En consecuencia, este estudio constituye una primera etapa de investigación, pudiendo ser ampliado y/o comparado con otras bases de datos, además de indagar sobre el perfil estadístico de los datos para permitir gestar conclusiones con otras aristas.

7. Referencias bibliográficas:

Benford, F. 1938. The law of anomalous numbers. *Proceedings of the American Philosophical Society*. 78(4):551-572.

Burke, J. and E. Kincanon (1991). Benfordís law and physical constants: the distribution of initial digits. *American Journal of Physics* 59, 952

Carslaw, C. A. P. N. 1988. Anomalies in income numbers: Evidence of goal oriented behavior. *The Accounting Review*. LXIII(2):321-327.

Castañeda, G. 2011. La ley de Benford y su aplicabilidad en el análisis forense de resultados electorales. *Scielo. Política y Gobierno*. Vol. 18 nº 2 Mexico

Etteridge M. L. and R. P. Srivastava. 1999. Using digital analysis to enhance data integrity. *Issues in Accounting Education*. 14(4):675-690.

Furlan, L. 1948. Das Harmoniegesetz der Statistik: Eine Untersuchung uber die metrische Interdependenz der sozialen Erscheinungen, Basel, Switzerland -G xiii:504.

Golbeck J, 2015 “Benford’s Law Applies to Online Social Networks,”PLOS ONE, v. 10, nº 8

Hill, T. P. 1995. A statistical derivation of the significant digit law. *Statistical Science.* 10(4):354-363.

Informe 16 de Federación Argentina de Consejo Profesionales de Ciencias Económicas (FACPCE). 2009. “Ríesgo de Auditoría y Significación” pàg. 31 a 52.

Newcomb, S. 1881. Note of the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*. 4:39-40.

Nigrini, M. J. 1996. Taxpayer compliance application of Benford’s law. *Journal of the American Taxation Association*. 18(1):72-92.

Nigrini, M. J. and L. J. Mittermaier. 1997. The use of Benford’s law as an aid in analytical procedures. *Auditing: A Journal of Practice & Theory*. 16(2):52-67.

Nigrini, M. J. 1999. Adding value with digital analysis. *The Internal Auditor*. 56(1):21-23.

Nigrini, Mark and Miller Steven J., 2007. Benford’s Law Applied to Hydrology Data—Results and Relevance to Other Geophysical Data. *International Association for Mathematical Geology.* Math Geol (2007) 39: 469–490

Nigrini, Mark Benford’s Law: Applications for forensic accounting, auditing, and fraud detection, vol. 586. John Wiley & Sons, 2012

Normas Internacionales de Auditoría (NIA) (2014) 300 a 500.

Pepijn de Vries, Albertinaka J. 2013 Compliance of LC50 and NOEC data with Benford's Law: An indication of reliability? Elsevier Ecotoxicology and Environmental Safety 201 3.

Roukema, B. F. “Benford’s Law anomalies in the 2009 Iranian presidential election,” Unpublished manuscript, 2009.

Sambridge M., Hrvoje T.,Arroucau P (2011) “Benford’s Law of First Digits: From Mathematical Curiosity to Change Detector”, Asia Pacific Mathematics Newsletter

Thomas, J. K. 1989. Unusual patterns in reported earnings. *The Accounting Review*. LXIV(4):773-787.

Varian, H. R. 1972. Benford's law. *The American Statistician*. 26:65-66.

Wallace, W. A. 2002. Assessing the quality of data used for benchmarking and decision-making. *The Journal of Government Financial Management.* (Fall) 51 (3):6-22.