

"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"

Lic. H. Rubén MORALES, Dra. Cecilia B. DÍAZ, Dr. Ricardo J. CASTELLO

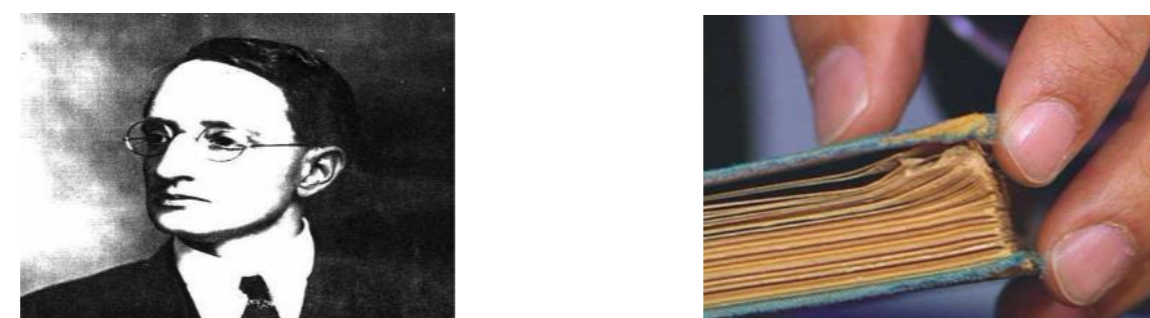
Introducción

La Ley de Benford considera que en un conjunto determinado de números, más del 30% 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. La finalidad de este trabajo es verificar si la distribución estadística de Benford se aplica a los números representados por el cantidad de registros (tamaño) 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.

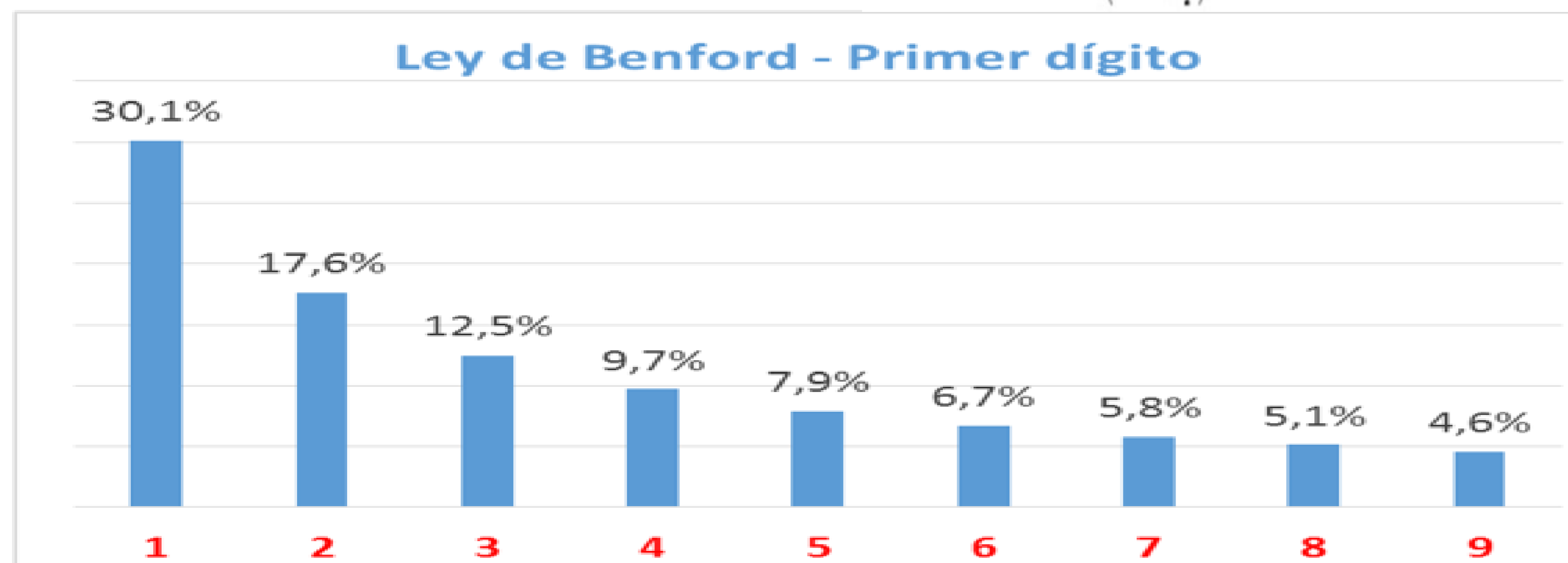
Marco Teórico:

La Ley de Benford



Benford (1938) advirtió que su libro de tablas de logaritmos tenía más desgastadas las primeras páginas que las últimas. Dedujo que usó más números con cifras iniciales bajas (1, 2 o 3) y menos con las que empiezan con dígitos altos (7, 8 o 9). Concluyó que la distribución de la primera cifra de los números no es uniforme (1/9). Experimentó con 20,229 números de muestras distintas: longitud de ríos, cantidad de habitantes de ciudades, constantes físicas y químicas (como peso atómico de los elementos), entre otras. De los resultados postuló que la probabilidad de que el primer dígito sea d, responde a la siguiente función logarítmica y distribución porcentual:

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



AÑO	AUTOR	APLICACIÓN LEY DE BENFORD A OTRAS DISCIPLINAS
1972	Variam	Datos bursátiles
1991	Burke y Kincanon	Constantes Matemáticas
1996	Nigrini	Contabilidad – Fraudes contables
2009	Roukema	Fraudes Electorales en Irán
2011	Sambridge	Estadística y previsión de terremotos
2013	Pepijn de Vries	Indicador para evaluar riesgo de toxicidad
2015	Golbeck	Cantidad de seguidores en redes sociales

AÑO	AUTOR	OPINIÓN SOBRE LA LEY DE BENFORD
1942	Furlan	"Es la verdad de la naturaleza"
1972	Variam	"Puede servir de prueba de honestidad o validez de datos"
1995	Hill	"Fenómeno empírico, como lo es la distribución normal"
1996	Nigrini	"La falta de cumplimiento puede denotar fraude"
1999	Etteridge	"No fraude, sino ineficiencias operativas o fallas"

Objetivos:

- Verificar si la distribución de Benford se aplica al conjunto de números dados por el tamaño (cantidad de registros) de las distintas tablas que conforman una BD.
- Determinar su uso como posible indicador de riesgo inherente (o que viene dado) de la información que recibe el auditor para su control.

Metodología:

Para verificar empíricamente si la ley de Benford se ajusta a la cantidad de registros de las tablas de una base de datos, se recurrió a datos de un back up de una empresa que opera en el sector energético. La BD es relacional bajo Oracle, con distintos módulos funcionales. La BD se conformó de 8 módulos heterogéneos, en cuanto al tipo de procesos como al volumen de la información que almacenan. Ello generará una mayor robustez en los resultados a que se arriben.

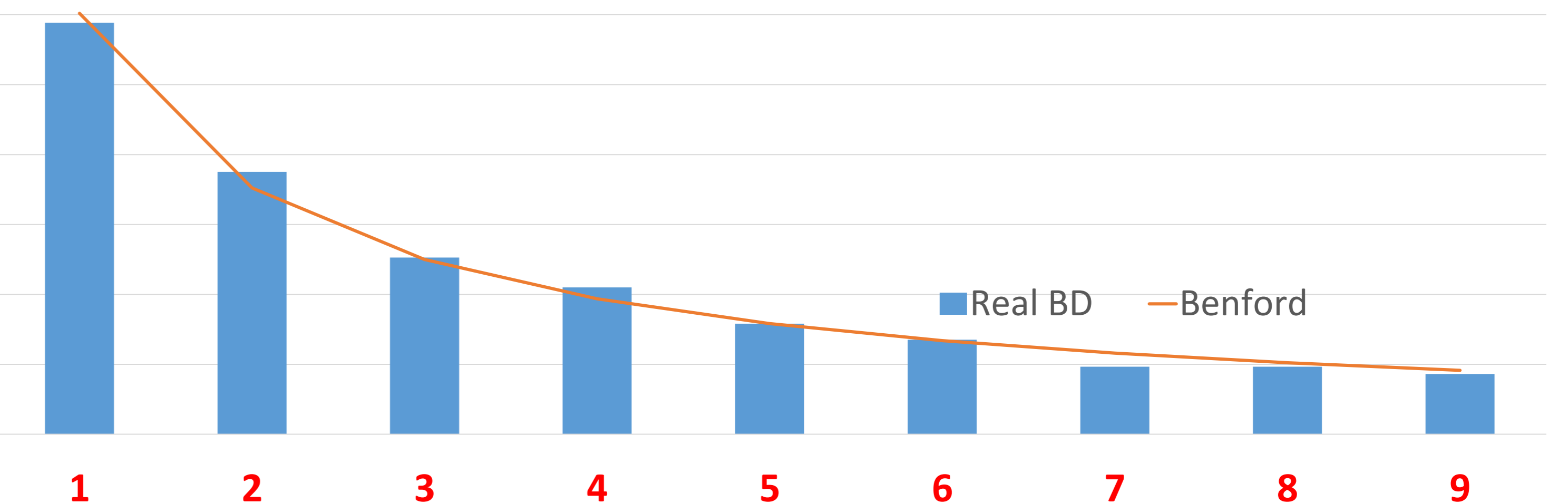
Módulos informáticos que forman la Base de Datos

Módulo informático	Cantidad tablas	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
Adm. Expedientes (EX)	100	13.416.811
Total Base de Datos	1.923	4.485.813.693

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%

Resultado: Análisis Primer Dígito del tamaño de las tablas de BD



Pruebas de bondad de ajuste:

Prueba χ^2 (Chi cuadrado)									
	CO	RH	IN	GO	SI	GC	SU	EX	Total
χ^2	6,09	9,48	6,15	10,98	5,17	8,55	11,34	6,51	6,77

Prueba χ^2 (Chi cuadrado): < 15,51 Acepta --> cumple distribución de Benford

Prueba Desviación Media Absoluta									
	CO	RH	IN	GO	SI	GC	SU	EX	Total
Tablas	78	340	71	58	87	808	381	100	1.923
Tamaño	225	1,5	15	0,4	1,8	4.093	135	13	4.485
MAD	0,0213	0,0172	0,0251	0,0434	0,0203	0,0102	0,0130	0,0238	0,0049
Califica	Bajo	Bajo	Bajo	Bajo	Bajo	Acepta	Medio	Bajo	ALTO

Motivo "Bajo": reducida cantidad de tablas y/o datos en los módulos

(*) en millones de registros

Conclusiones

	Verifica Benford?	Verifica χ^2	Verifica MAD	Causa	Solución Posible
Base Datos	SI	SI	SI		
Módulos	SI	SI	PARCIAL	Pocas tablas o datos	Agregar Back up - Análisis temporal

Propuestas:

- 1- Indagar la factibilidad de fijar un orden de prioridad entre los módulos, donde se debe orientar la atención de la auditoría o área de control.
- 2- Indicador para el auditor de la confianza o alerta del posible riesgo inherente o preexistente de los datos informatizados

Referencias bibliográficas:

- Benford, F. 1938. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4):551-572.
- Burke, J. and Kincanon 1991. Benford's law and physical constants: the distribution of initial digits. *American Journal of Physics* 59, 952.
- Carlsaw, C. 1988. Anomalies in income numbers: Evidence of goal oriented behavior. *The Accounting Review*, LXII(2):321-327.
- Castañeda, G. 2011. La ley de Benford y su aplicabilidad en el análisis forense de resultados electorales. *SciELO. Política y Gobierno*, V 18 n° 2 Mexico.
- Etteridge M. L. and R. P. Srivastava. 1999. Using digital analysis to enhance data integrity. *Issues in Accounting Education*, 14:675-690.
- Furlan, L. 1948. Das Harmoniegesetz der Statistik: Eine Untersuchung über 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, n° 8
- Hill, T. 1995. A statistical derivation of the significant digit law. *Statistical Science*, 10:354-63.
- Informe 16 de Fed. Argentina de Consejo Prof. Cs. Económicas (FACPCE). 2009. "Riesgo de Auditoría y Significación" pág. 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, M. and Miller S. 2007. Benford's Law Applied to Hydrology Data, Results and Relevance to Other Geophysical Data. *International Assoc. 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? *Elsivier Ecotoxicology and Environmental Safety* 2013.
- 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*