



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO

---

---

FACULTAD DE CIENCIAS

Propuesta Metodológica para Utilizar la Ley  
de Benford en la Investigación Forense en  
Resultados Electorales

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuario

PRESENTA:

Óscar Jiménez Sánchez

TUTOR

Dr. Vicente Torres Zúñiga

Ciudad Universitaria, CD. MX. 2020





Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Hoja de datos del jurado

### **1. Datos del alumno**

Óscar Jiménez Sánchez  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Actuaría  
No. Cuenta: 304200706

### **2. Datos del tutor**

Dr. Vicente Torres Zúñiga

### **3. Datos del sinodal 1**

M. en A. Blanca Dulce Miriam Benítez Pérez

### **4. Datos del sinodal 2**

Act. Hortensia Pescador Romo

### **5. Datos del sinodal 3**

M. en I. Saulo Gonzalo Carmona Contreras

### **6. Datos del sinodal 4**

Fis. Jimmy Hernández Morales

### **7. Datos del trabajo escrito**

Título: Propuesta metodológica para utilizar la ley de Benford en la investigación forense en resultados electorales.

Páginas: 102

# Agradecimientos

Antes que nada quiero agradecer a la UNAM. por haberme permitido estudiar y desarrollarme en ésta gran institución. A mi mamá por toda la entrega y dedicación que le dio a mi vida con el objetivo de llegar a éste momento. Este trabajo se lo dedico a ella como símbolo de todo el esfuerzo que hizo por mi para sacarme adelante aún en los tiempos más cruciales. Al Mtro. José Guadalupe Bañuelos Muñeton por su colaboración en la revisión técnica de este proyecto de tesis, sus comentarios y observaciones fueron muy importantes y fundamentales para poder avanzar. A mi asesor que tuvo la paciencia y dedicación en éste trabajo. A mi hermana por haberme apoyado en momentos más difíciles. A mi tío que estuvo con nosotros en momentos críticos. Y a mis verdaderos amigos que me sacaron una risa y que estuvieron conmigo en las peores situaciones.

Finalmente, debo resaltar que este trabajo ha recibido el apoyo económico de una entidad pública, la Universidad Nacional Autónoma de México, mediante el proyecto: UNAM-PAPIME-PE115519.

Esta tesis se redactó y maquetó en L<sup>A</sup>T<sub>E</sub>X.

# Resumen

En la presente tesis, realizamos una exploración de la ley de Benford para encontrar anomalías en bases de datos con resultados electorales. En la literatura académica se sugiere ampliamente utilizar la ley de Benford para identificar alteraciones en libros contables, fiscales, así como en registros de votos. Sin embargo, mucha de la metodología fundamental sugerida en tales obras puede mejorar su objetividad, presentación e interpretación de resultados. Con el fin de que la auditoría forense sea más inteligible para los profesionales del Derecho, las autoridades correspondientes (jueces, tribunales, abogados, etcétera.), y para el público lego.

Proponemos utilizar la sucesión de Fibonacci (por ser un ente matemático objetivo con cierta relación con fenómenos naturales, como puede ser la proporción espiral de una concha marina) como base explicativa para observar el comportamiento de la ley de Benford y homologar los índices estadísticos que suelen presentar los peritos del área. De modo que la presentación e interpretación que reciben las autoridades (jueces y tribunales) sea más clara. Para lograr tal objetivo, en el límite de concordancia entre la serie de Fibonacci y la ley de Benford realizamos una manipulación intencional de la información. Entonces, contamos una base de datos fiel a la ley de Benford y controlamos su degradación hasta cierto umbral. En tal condición observamos cuales son los valores de seis índices estadísticos: **1)** Desviación Media Absoluta (por sus siglas en inglés: MAD), **2)** Correlación, **4)** Concordancia por modelo lineal, **5)** Prueba  $\chi^2$  y **6)** Prueba-Z. Tales índices presentan ventajas y desventajas para el quehacer forense, las cuales comentamos a lo largo de este trabajo.

Para mostrar cómo puede funcionar un estudio empírico, probamos la ley de Benford con tres bases de datos electorales, correspondientes a las elecciones mexicanas presidenciales de los años 2006 y 2018; además, incluimos las elecciones presidenciales de España 2015.

Si bien muchas clases de bases de datos siguen la ley de Benford y se puede argumentar la alteración de datos cuando no la siguen. Es importante establecer el límite y alcances de tal técnica. Para ello, en este trabajo de tesis realizamos una exposición de esta técnica para el operador de campo,

brindándole un sustento metodológico. De hecho, brindándole herramientas de análisis computacional.

Finalmente, en este contexto de aplicar la ley de Benford contra el fraude electoral, es importante señalar que si las anomalías sobresalen al superar índices estadísticos, en lugar de declarar la existencia de un delito, aquí se sugiere realizar una investigación tanto dentro de la base de datos como en el proceso electoral con el fin de profundizar en los indicios de una posible manipulación maliciosa de los datos. De modo que la ley de Benford es una prueba presuntiva, una especie de semáforo metafórico de una parte de la limpieza de la elección. Esta perspectiva de dictaminación es común en periciales e investigación forense.

# Índice general

|   |            |
|---|------------|
| <b>Resumen</b>  | <b>II</b>  |
| <b>Índice de figuras</b>                                      | <b>VII</b> |
| <b>Índice de tablas</b>                                       | <b>IX</b>  |
| <b>1. Introducción</b>  | <b>3</b>   |
| 1.1. Contexto histórico . . . . .                             | 4          |
| 1.2. Sucinta descripción de la distribución Benford . . . . . | 7          |
| 1.2.1. Extensión a más cifras . . . . .                       | 11         |
| 1.3. Explicación del mecanismo de la ley de Benford . . . . . | 11         |
| 1.4. Una deducción matemática de la ley de Benford . . . . .  | 14         |
| 1.5. Validez en otras bases numéricas . . . . .               | 16         |
| 1.6. Limitaciones de la ley de Benford . . . . .              | 17         |
| 1.6.1. Otras limitaciones de la ley de Benford . . . . .      | 18         |
| 1.7. Conclusiones del capítulo . . . . .                      | 18         |
| <b>2. Breviario de aplicaciones de la Ley de Benford</b>      | <b>20</b>  |
| 2.1. Área físico-matemática . . . . .                         | 20         |
| 2.1.1. Detección de imágenes manipuladas . . . . .            | 23         |
| 2.1.2. Resolución de preguntas de opción múltiple . . . . .   | 24         |
| 2.2. Área económica . . . . .                                 | 24         |
| 2.2.1. Aplicaciones en auditorías . . . . .                   | 26         |
| 2.2.2. Evasión de impuestos . . . . .                         | 27         |
| 2.3. Área electoral . . . . .                                 | 28         |
| 2.3.1. Caso: Elecciones mexicanas-2006 . . . . .              | 29         |
| 2.4. La estadística en el contexto forense . . . . .          | 32         |
| 2.5. Conclusiones del capítulo . . . . .                      | 34         |

---

|  |           |
|--|-----------|
| <b>3. Análisis estadístico para identificar irregularidades electorales</b>          | <b>35</b> |
| 3.1. Malas prácticas electorales . . . . .   | 35        |
| 3.2. Alcances del análisis estadístico forense de elecciones . . . . .               | 38        |
| 3.3. Consideraciones estadísticas . . . . .  | 39        |
| 3.4. Problemas de multiplicidad estadística . . . . .                                | 41        |
| 3.5. Análisis de distribuciones numéricas de dígitos . . . . .                       | 43        |
| 3.6. Prueba de hipótesis nula basada en dígitos . . . . .                            | 44        |
| 3.7. Entre la precisión y la irregularidad electoral . . . . .                       | 46        |
| 3.8. Conclusiones del capítulo . . . . .   | 47        |
| <b>4. Propuesta de análisis y auditoría</b>  | <b>50</b> |
| 4.1. Criterios de ajuste entre modelos teóricos y datos empíricos . .                | 51        |
| 4.1.1. Desviación media absoluta MAD . . . . .                                       | 51        |
| 4.1.1.1. Interpretación de MAD . . . . .   | 52        |
| 4.1.2. Prueba ji-cuadrada . . . . .  | 53        |
| 4.1.3. Prueba-Z . . . . .  | 56        |
| 4.1.4. Prueba-Z para diferencia de proporciones . . . . .                            | 57        |
| 4.1.5. Análisis por regresión lineal . . . . .                                       | 57        |
| 4.2. Cuarteto de Anscombe . . . . .  | 60        |
| 4.3. Sucesión de Fibonacci . . . . .   | 60        |
| 4.4. Calibración de índices estadísticos a partir de la serie de Fibonacci . . . . . | 62        |
| 4.5. Conclusiones del capítulo . . . . .   | 68        |
| <b>5. Análisis electorales del primer dígito</b>                                     | <b>69</b> |
| 5.1. Elecciones presidenciales de México 2006 . . . . .                              | 69        |
| 5.2. Elecciones España 2015 . . . . .  | 73        |
| 5.3. Elecciones presidenciales México 2018 . . . . .                                 | 76        |
| 5.4. Conclusiones de capítulo . . . . .  | 81        |
| <b>6. Conclusiones generales</b>   | <b>82</b> |
| 6.1. Prospectiva de la investigación . . . . .                                       | 83        |
| 6.2. Palabras finales . . . . .  | 83        |
| <b>Bibliografía</b>  | <b>84</b> |
| <b>Anexos</b>  | <b>87</b> |

|   |           |
|---|-----------|
| <b>A. Apéndice: Códigos octave</b>                              | <b>88</b> |
| A.1. Disección de los códigos . . . . .                         | 88        |
| A.1.1. Programa de la moneda . . . . .                          | 88        |
| A.1.2. Los programas adoptados por Marco Cococcioni . . . . .   | 89        |
| A.1.3. Código de sustitución de primeros dígitos y análisis MAD | 93        |
| <b>B. Presentaciones en eventos académicos</b>                  | <b>95</b> |
| <b>C. Tablas logarítmicas de Newcomb</b>                        | <b>99</b> |

# Índice de figuras

|      |  |    |
|------|--|----|
| 1.1. | Fotografía de Simon Newcomb. . . . .   | 4  |
| 1.2. | Fotografía de Frank Albert Benford, Jr. . . . .  | 5  |
| 1.3. | Fotografía de Mark Nigrini . . . . .   | 6  |
| 1.4. | Sitio Web donde se pueden visualizar el ajuste de bases de datos a la ley de Benford. . . . .  | 7  |
| 1.5. | Gráficas de la probabilidad de aparición del primer al cuarto dígito . . . . .   | 10 |
| 1.6. | Tres distribuciones que no se ajustan a la ley de Benford, hasta que son combinadas . . . . .  | 11 |
| 1.7. | Probabilidades porcentuales de la aparición del segundo dígito en sistema decimal. . . . .   | 12 |
| 1.8. | Probabilidades aparición de primer dígito en diferentes sistemas [20]. . . . .   | 13 |
| 1.9. | Probabilidades aparición de primer dígito en diferentes sistemas. . . . .  | 16 |
| 2.1. | Gráficas que que utilizan datos físicos y matemáticos . . . . .  | 21 |
| 2.2. | Retrato Walter R. Mebane jr . . . . .  | 28 |
| 2.3. | Captura de pantalla del sitio llamado «Análisis Forense de Procesos Electorales», último acceso 28/01/2020. . . . .  | 30 |
| 2.4. | Ley de Benford contrastado datos electorales del 2006 . . . . .  | 32 |
| 4.1. | Esquema de los elementos de la prueba Z, la curva (línea negra) representa la función de densidad y el área bajo la curva (color azul) representa la probabilidad acumulada. . . . . | 56 |
| 4.2. | Interpretación para $R^2$ en función de las variaciones de datos . . . . .   | 59 |
| 4.3. | Gráficas del cuarteto de Anscombe . . . . .  | 61 |
| 4.4. | Gráfica y ajuste exponencial a la sucesión de Fibonacci hasta $f_{20}$ . . . . .   | 62 |
| 4.5. | Los primeros 152 valores de la sucesión de Fibonacci, graficados en escala logarítmica. . . . .  | 63 |
| 4.6. | Ilustración del método de sustitución de 1 a 9 utilizado. . . . .  | 64 |

---

|      |   |    |
|------|---|----|
| 4.7. | Sustitución de los primeros dígitos, 1 por 9, de los primeros 175 valores de la sucesión de Fibonacci . . . . .                                     | 65 |
| 4.8. | Gráficas de regresiones lineales de datos de la sucesión de Fibonacci para 152 elementos y 175 elementos con sustitución . . . . .                  | 67 |
| 5.1. | Linealización de los datos de los dos candidatos con más votos en la elección presidencial México 2006. . . . .                                     | 73 |
| 5.2. | Datos de la elección presidencial de España 2015 linealizados con la ley de Benford. . . . .  | 75 |
| 5.3. | Datos de la elección presidencial de México 2018 linealizados con la ley de Benford. Solo se muestran ajustes para tres partidos políticos. . . . . | 80 |
| B.1. | Constancia de participación el tercer foro de avances estudiantiles, en la Red Temática CONACyT de Ciencias Forenses, C.U., 11/sept/2018. . . . .   | 95 |
| B.2. | Cartel presentado en el tercer foro de avances estudiantiles, en la Red Temática CONACyT de Ciencias Forenses, C.U., 11/sept/2018. . . . .          | 96 |
| B.3. | Constancia de participación el primer simposio estudiantil de estadística forense. . . . .  | 97 |

# Índice de tablas

|  |    |
|--|----|
| 1.1. Distribuciones de la ley de Benford para la $k$ -ésima posición inicial calculado. . . . .  | 9  |
| 2.1. Distribución porcentual de los primeros dígitos de datos físicos y matemáticos . . . . .  | 22 |
| 2.2. Datos de primer-dígito aplicados a datos macroeconómicos [29].  | 25 |
| 4.1. Criterios de interpretación de MAD, de acuerdo con M. Nigrini.  | 52 |
| 4.2. Mejora del criterio MAD aplicado a la sucesión de Fibonacci en función del número de elementos de la sucesión $N$ . . . . .   | 63 |
| 4.3. Índices estadísticos de los datos Fibonacci-152 y Fibonacci-175 sustituido. Se muestra el criterio de alarma para Fibonacci-175 sustituido . . . . .  | 67 |
| 5.1. Valores de las linealizaciones y MAD de los datos electorales de los partidos que participaron en la contienda presidencial de México 2006. En tipo negritas están los indicadores que no cumplieron el criterio de umbral propuesto. . . . . | 70 |
| 5.2. Valores de índices estadísticos de los datos electorales de los partidos que participaron en la contienda de España 2015. . .   | 74 |
| 5.3. Valores de las linealizaciones y MAD de los datos electorales México 2018. . . . .  | 77 |

# Preámbulo

## Estructura de la tesis

Esta obra se dividió en las siguientes secciones

- En el capítulo 1 explicamos el contexto histórico de la ley de Benford y algunas aplicaciones reportadas en la literatura académica general.
- En el capítulo 2 presentamos un marco teórico matemático breve de los fundamentos y propiedades de la ley Benford.
- En el capítulo 3 explicamos (de modo sucinto y didáctico) en que consisten seis índices estadísticos utilizados en este trabajo: **1)** Desviación Media Absoluta, **2)** Correlación, **4)** Concordancia por modelo lineal, **5)** Prueba  $\chi^2$  y **6)** Prueba-Z.
- En el capítulo 4 presentamos una homologación de análisis de bases de datos ante la ley de Benford, proponiendo la serie de Fibonacci como el estándar objetivo de comparación.
- El capítulo 5 muestra la concordancia de la ley de Benford y tres bases de datos electorales presidenciales: **1)** México-2006, **2)** España-2015, y **3)** México-2018.
- Finalmente, en el capítulo 6 presentamos nuestras conclusiones principales y mencionamos el trabajo a futuro que se requiere investigar.

Adicionalmente presentamos un apéndice con los códigos de las rutinas programas en *Octave* (un software gratuito para realizar cálculo numérico).

---

## Puntos a destacar:

A la consideración del autor, los tres principales aportes de este trabajo de tesis son

- Una exposición amplia y didáctica de la ley de Benford accesible a diferentes perfiles profesionales, con especial énfasis para ser comprendida por actuarios, científicos forenses y profesionales del derecho.
- Plantea un criterio objetivo y homologado, posiblemente estándar, para realizar un análisis de la concordancia entre la ley de Benford y una base de datos.
- Como evidencia empírica, muestra como tres diferentes bases de datos siguen –en cierto grado– la tendencia de la ley de Benford, especulando sobre su reputación histórica.

## Palabras clave:

1) Ley de Benford 2) Investigación forense 3) Invarianza de escala 4) Ley de potencias 5) Distribución del primer dígito.

# Capítulo 1

## Introducción

Sabemos que hoy en día, el sistema decimal es el conjunto que más empleamos para numerar objetos, para medir, contar o etiquetar en nuestra cotidianidad.

Parece obvio esperar que el primer dígito de números en un conjunto de datos de fuentes diversas, de magnitudes variadas (disimiles y estocásticos) que aparecen en la primera página de un periódico de economía, presenten una probabilidad igual. Puesto que la teoría de la probabilidad nos muestra que la probabilidad de un evento cualquiera deber ser igual al número de casos favorables sobre casos totales. Y supone también que todos los eventos son equiprobables, tal es el ejemplo clásico de un dado teórico perfectamente balanceado. El evento de que salga una cara del dado ideal debe ser igual a  $1/6$ . Para la aparición del primer dígito en una base de datos como puede ser una electoral, sabemos que el primer dígito no puede ser cero, por tanto, la probabilidad para cualquier otro dígito debería ser  $1/9$ . Pero la realidad es otra, la probabilidad de encontrar el dígito 1 es mayor, que la del dígito 2, y disminuye con el 3 y así continua, de modo que el dígito 9 es el que presenta la mínima probabilidad de iniciar un número que es parte de un registro de base de datos.

A continuación, presentamos un panorama histórico del descubrimiento, evaluación y aplicación de esta relación matemática.

---

## 1.1. Contexto histórico

El logaritmo es una herramienta matemática que facilita que una operación de potencias se convierta en una multiplicación, y las multiplicaciones en sumas. Por sus ventajas en simplificar los cálculos, cuando se descubrieron en el siglo XVII, fueron rápidamente adoptados por científicos, ingenieros, banqueros, además de otros profesionales; quienes entre sus instrumentos cotidianos de trabajo utilizaban tablas con los valores logarítmicos.

En 1881, el matemático y astrónomo Simón Newcomb se percató que las tablas logarítmicas, –impresas en libros disponibles en bibliotecas– mostraban mayor desgaste en las primeras páginas que en las últimas. Es decir, que las hojas relacionadas con los primeros dígitos (*e.i.* 1, 2, 3) se encontraban en peores condiciones que los folios de los últimos dígitos (*i.e.* 7, 8, 9). Sin importar el campo de uso, el mismo patrón de deterioro lo mostraba el libro consultado por el astrónomo que el contador. Por este simple hecho, él dedujo que los dígitos iniciales muestran una diferente probabilidad de ocurrencia comparativa. En otras palabras, él conjeturó que el dígito 1 aparece con mayor frecuencia, seguido del 2, hasta el 9 que es el menos frecuente.

De este razonamiento Newcomb manifestó verbalmente una relación o ley logarítmica, que se puede enunciar en términos modernos como: «la ley de probabilidad de la ocurrencia de números es tal que las mantisas de sus logaritmos son equiprobables». Sin embargo, por mucho tiempo esta observación sólo fue una curiosidad matemática para algunos especialistas en la materia.

En 1938, el físico e ingeniero Frank Albert Benford Jr. (quien trabajó para la compañía *General-Electric*) observó el mismo fenómeno en las tablas de

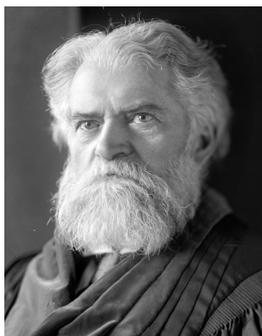


Figura 1.1: Fotografía (tomada entre 1905 y 1909) del astrónomo, matemático y economista Simon Newcomb (1835-1909), a quién se le atribuye el descubrimiento de la variación de frecuencias en la aparición de los primeros dígitos en bases de datos contables, que después se conocería como ley de Benford.

---

logaritmos; y emprendió un registro ingente de datos. Él obtuvo 20,229 números agrupados en 20 muestras de gran diversidad, entre las que se encontraban: áreas fluviales, constantes, magnitudes físicas y químicas, funciones matemáticas, poblaciones en ciudades, datos financieros, promedios de la liga americana de baseball, incluso números de direcciones de personas. A partir de tales resultados empíricos Benford postuló una «ley de los números anómalos». Sin embargo, Benford, sólo alcanzó a describir la relación, sin profundizar por las razones, mecanismos o causas de tal comportamiento, por lo que por mucho tiempo sólo se contaba con esta teoría de caja negra.

En 1961, el matemático Roger Pinkham aportó una explicación a esta relación, su reflexión se basó en que realmente existe una ley de frecuencias de dígitos que debe ser universal. Tanto si se calcula los precios entre distintas divisas, o si se mide la longitud entre diferentes sistemas métricos; en todos estos casos, las proporciones de frecuencias de dígitos deberían ser las mismas. Así, la distribución de las frecuencias de dígitos debía ser invariante frente a cambios de escala; y que esta constituía a la ley de Benford. Boyle, en 1994, demostró que si los datos analizados se construyen a partir de multiplicaciones y cocientes de variables aleatorias provenientes de diversas fuentes, también satisfacen la antes mencionada ley [1].

En 1995, el matemático Theodore P. Hill proporcionó una demostración matemática más formal. La demostración argumenta en base de algunos teoremas del límite central y su relación con el comportamiento de las mantisas en las multiplicaciones de valores aleatorios. La mayoría de nuestro referente teórico matemático proviene de este punto en la historia.

Como hito histórico de las aplicaciones de la ley de Benford, debemos señalar 1994, cuando Mark Nigrini, profesor de contabilidad de Dallas, propuso



Figura 1.2: Fotografía de Frank Albert Benford, Jr., reconocido como el redescubridor y por generalizar la ley que lleva su nombre, un postulado estadístico sobre la frecuencia de aparición de dígitos en la primera posición de los datos de diversas listas.

---

emplear el análisis de las frecuencias de los dígitos como mecanismo analítico para detectar posibles situaciones de fraude e irregularidades. Su impulso (hasta la fecha) ha popularizado la ley de Benford. Inicialmente lo aplicó al estudio de datos fiscales y programó un algoritmo para detectar en que porcentaje algunos datos contables encajan con la ley de Benford. Desde entonces, la ley de Benford ha sido parte de discusiones académicas acaloradas. Por ejemplo, en situaciones para detectar fraude electoral [2]-[7], identificar errores de redondeo, manipulación en libros contables [8]-[10], detección de anomalías en series de datos físicos [11]-[15], o bien como estrategia para resolver adecuadamente preguntas de opción múltiple en exámenes [16]. Detección de fraude financiero, científico, o bien en transcripción del genoma. De hecho, de modo metafórico de uróboros (una serpiente que engulle su propia cola), las cifras de artículos que citan a la ley de Benford, siguen la ley de primer dígito [17]. El tema de aplicaciones de la ley de Benford nos parece relevante, por lo que lo desarrollamos a detalle en las siguientes secciones.

Con todo, en México es poco conocida como medio de prueba en situaciones donde se requiere un análisis forense. Por ejemplo, si bien esta ley se enseña en la carrera técnica de contaduría del Colegio Nacional de Educación Profesional Técnica (CONALEP), no aparece en sus temarios de asignaturas. Por su parte, en la Universidad Nacional Autónoma de México UNAM, hasta mayo del 2018 encontramos sólo una tesis que contenía el título de Benford, pero su campo de aplicación fue la física estadística [18], no encontramos trabajos en el área de la actuaría u afines: economía, contaduría, ciencia forense, etcétera. Es posible que muchas personas consideren a la ley de Benford como una curiosidad matemática, en lugar de una herramienta de análisis. Lo que explicaría que existen páginas web sobre el tema (ver la Fig. 1.4), pero



Figura 1.3: Fotografía de Mark Nigrini, uno de los mayores impulsores del uso de la ley de Benford para encontrar fraudes en libros contables y otras bases de datos.

pocos trabajos académicos en instituciones tan grandes y prestigiadas como la UNAM.

## 1.2. Sucinta descripción de la distribución Benford

¿Pero qué es la ley de Benford?, ¿cómo se llega a dicha ley?

Al observar las tablas logarítmicas, Albert Frank Benford observó que al calcular el porcentaje promedio de veces en los que aparece el número 1 como primer dígito de dichas tablas, se dio cuenta que éste número aparece con más frecuencia que el número 9, con una frecuencia promedio de 30.6% de veces, cuya aproximación es  $\log(2)$ . Luego se percató que el número 2 aparece con una frecuencia de 18.5% aproximadamente a  $\log(3/2) = \log(3) - \log(2)$ .

Siguiendo el razonamiento llegó al otro extremo de la distribución, encontró que el número 9 aparece con frecuencia en proporción de 4.7% cuya aproximación es  $\log(10/9)$  o que es lo mismo  $\log(10) - \log(9)$ . A toda esta relación matemática Benford infirió que la ocurrencia de éstos números sugieren una distribución logarítmica.

La distribución teórica para el dígito ubicado en la  $i$ -ésima posición (de



Figura 1.4: Captura de pantalla del sitio Web donde se pueden visualizar el ajuste de bases de datos a la ley de Benford. Dirección web:<http://testingbenfordslaw.com/> (último acceso: 21/01/2020).

---

izquierda a derecha) de números generados de acuerdo con determinado proceso estadístico. En particular, la frecuencia relativa  $P$  que caracteriza a la ley de Benford para el primer dígito se describe como:

$$P(d_i) = \log_{10} \left( 1 + \frac{1}{d_i} \right); \quad d_i \in \{1, 2, 3, \dots, 9\}. \quad (1.1)$$

de este modo, el dígito 1 muestra una probabilidad de 0.301 mientras que el dígito 9 muestra una probabilidad de tan sólo 0.0458. Las probabilidades restantes en un cuadro más adelante, donde se señalan las distribuciones de la ley de Benford para diferentes dígitos.

Ahora, la distribución teórica que caracteriza la ley de Benford para el segundo dígito se describe mediante la siguiente expresión:

$$P(d_2) = \sum_{k=1}^9 \log_{10} \left( 1 + \frac{1}{10k + d_2} \right). \quad (1.2)$$

[2]

En este caso, si existe una probabilidad positiva de que exista el dígito 0, que es igual a 0.11968, pues en este caso el cero sí puede presentarse en las segunda posición. La tabla 1.1 presenta el calculo de los dígitos en diferentes posiciones en un número, desde la uno a la cinco. Destacamos que para posiciones posteriores a la tercer se requieren varias cifras significativas para distinguir los valores de probabilidad. Por ejemplo, la posición quinta al presentar únicamente cuatro cifras significativas presenta una distribución uniforme causada por el redondeo. La Fig. 1.5 ilustra la idea.

Finalmente, esta ley de dígito-significativo se puede generalizar en términos de una densidad conjunta de dígitos en las primeras  $k$  posiciones iniciales, por ejemplo de la siguiente manera[2].

$$P(D_1 = d_1, \dots, D_k = d_k) = \log_{10} \left[ 1 + \left( \sum_{i=1}^k d_i \times 10^{k-i} \right)^{-1} \right] \quad (1.3)$$

Por ejemplo, calculemos la probabilidad relativa de 129 sean los primeros dígitos.[2] Esto es:

$$\begin{aligned} P(D_1 = 1, D_2 = 2, D_3 = 9) &= \log_{10} \left[ 1 + (1 \times 10^{3-1} + 2 \times 10^{2-1} + 9 \times 10^{1-1})^{-1} \right] \\ &= \log_{10} \left[ 1 + (129)^{-1} \right] \\ &\approx 0.00335. \end{aligned}$$

---

También se puede demostrar que la distribución del  $k$ -ésimo dígito significativo ( $D_k$ ) se aproxima muy rápido a una distribución uniforme cuando la  $k$ -ésima posición se mueve hacia la derecha [2].

Por medio de histogramas es sencillo visualizar la presencia de distribuciones no-uniformes para la primera y la segunda posición pero no para las demás posiciones. De hecho cuando  $k \geq 3$  las distribuciones asociadas tienen un valor medio de alrededor de 4.5 y una varianza de 8.25, valores que coinciden con los de una distribución uniforme.

Una forma equivalente a la expresión general a la Ec. 1.3 es:

$$P(\text{mantisa} \leq \frac{t}{10}) = \log_{10}(t); \quad t \in [1, 2, \dots, 10). \quad (1.4)$$

Algunas propiedades interesantes de la distribución logarítmica, demostradas en la literatura especializada, son las siguientes: Si  $X_1$  y  $X_2$  son variables aleatorias que se ajustan a la ley de Benford, entonces el resultado de las siguientes operaciones matemáticas también siguen la ley de Benford:

- $X_1 + X_2$ ,
- $X_1 \cdot X_2$ ,
- $\lambda X_1$ , cuando  $\lambda > 0$ .

Tabla 1.1: Distribuciones de la ley de Benford para la  $k$ -ésima posición inicial calculados mediante la Ec. 1.2. Se observa que para la primera posición es notoria la diferencia de probabilidades entre dígitos, en contraste para posiciones superiores a la frecuencia de distribución es uniforme, lo que destacamos para  $d_5$  presentando sólo cuatro cifras después del cero.

| Dígito ( $d_k$ ) | $P(d_1)$ | $P(d_2)$ | $P(d_3)$ | $P(d_4)$ | $P(d_5)$ |
|------------------|----------|----------|----------|----------|----------|
| 0                | -        | 0.11968  | 0.10178  | 0.10018  | 0.1000   |
| 1                | 0.30103  | 0.11389  | 0.10138  | 0.10014  | 0.1000   |
| 2                | 0.17609  | 0.10882  | 0.10097  | 0.10010  | 0.1000   |
| 3                | 0.12494  | 0.10433  | 0.10057  | 0.10006  | 0.1000   |
| 4                | 0.09691  | 0.10031  | 0.10018  | 0.10002  | 0.1000   |
| 5                | 0.07918  | 0.09668  | 0.09979  | 0.09998  | 0.1000   |
| 6                | 0.06695  | 0.09337  | 0.09940  | 0.09994  | 0.1000   |
| 7                | 0.05799  | 0.09035  | 0.09902  | 0.09990  | 0.1000   |
| 8                | 0.05115  | 0.08757  | 0.09864  | 0.09986  | 0.1000   |
| 9                | 0.04576  | 0.08500  | 0.09827  | 0.09982  | 0.1000   |

- $X_1^\lambda$ , cuando  $\lambda > 0$ ,
- $1/X_1$ .

Por otro lado, supongamos que reunimos al azar un conjunto de bases de datos que intrínsecamente no siguen la ley de Benford. Por ejemplo, mediciones de calores específicos en materiales, valores de pesos atómicos y datos matemáticos. Al combinar las bases de datos se obtendrá una mucho mejor coherencia con la ley de Benford. Es decir, la mezcla de colecciones de datos suele tender mejor a la ley de Benford que las bases de datos por separado [19]. La Fig. 1.6 ilustra la idea. Este resultado no es solo producido por un aumento en la cantidad de datos, es principalmente consecuencia de la recombinación de información. Este resultado será importante para establecer una metodología de lo general (análisis nacional general) a lo particular (estudio estatal o distrital) para estudiar grandes volúmenes de conteos, como sucede en una elección presidencial.

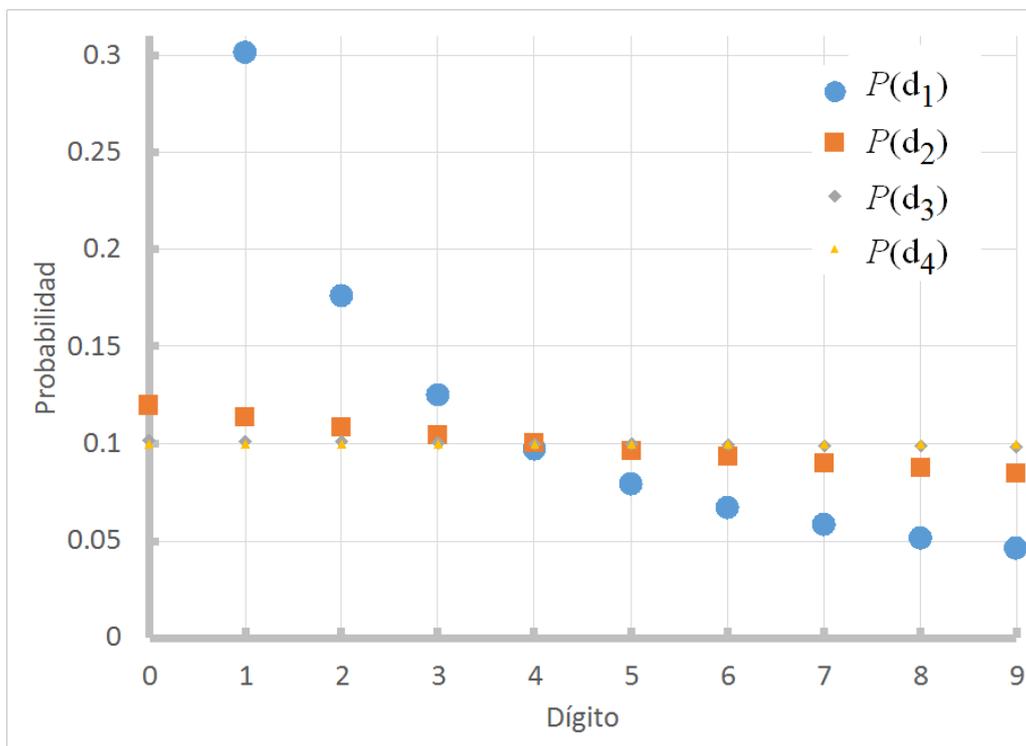


Figura 1.5: Gráficas de la probabilidad (utilizando los datos de la Tabla 1.1) de aparición del primer al cuarto dígito. Observe que para después de  $P(d_3)$  la distribución es prácticamente uniforme.

### 1.2.1. Extensión a más cifras

La ley se puede extender a más cifras. Cuando observamos la probabilidad de los primeros dos dígitos, por ejemplo: ¿Cuál es la probabilidad de que aparezca el 13 encabezando las entradas de una base de datos. Simplemente sustituimos  $d = 13$  de la ecuación (1.3).

La Fig. 1.7 muestra la probabilidad de que aparezcan los primeros dos dígitos en el sistema decimal. Aquí, el eje de las abscisas muestra los dos primeros dígitos, no el número *per se*, de modo que la posición de  $x = 24$  se vincula con la probabilidad de que los dígitos 2 y 4 aparezcan en tal secuencia u orden, y no de que aparezca el número 24 en sí mismo.

## 1.3. Explicación del mecanismo de la ley de Benford

A continuación, expondré una justificación simple sobre porque no es uniforme la frecuencia de aparición del primer dígito en un conteo. Esta explicación no pretende ser una demostración formal de la ley de Benford,

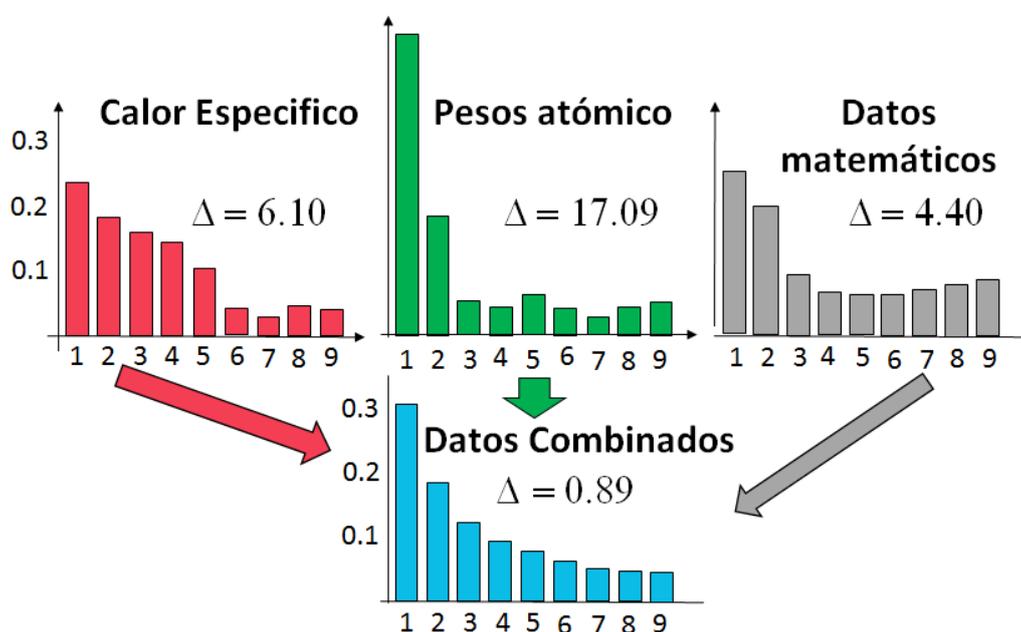


Figura 1.6: Los datos de distribuciones diferentes que originalmente no se ajustan a la ley de Benford, al combinarse sí la siguen. La figura muestra tres casos. La  $\Delta$  indica que tanto se separan los datos de la ley de Benford.

---

sólo un ejercicio para su comprensión.

Imaginemos que realizamos un conteo. Al inicio, lo hacemos del 1 al 9; por el momento todos los dígitos muestran la misma frecuencia de aparición. Al seguir el conteo, lo hacemos del 10 al 19; ahora el dígito 1 domina la distribución de frecuencias. Al continuar el conteo, en orden ganaran relevancia los siguientes dígitos. Cuando lleguemos al 99, todos los primeros dígitos mostraran la misma ocurrencia. Pero al continuar el conteo, tendremos que esperar un lapso mayor para obtener de nuevo la misma frecuencia uniforme. Es decir, números grandes derivados de un conteo tendrán mayor probabilidad de mostrar un dígito pequeño como primera cifra.

Ese fue un caso de conteo lineal, podemos sofisticar más el modelo. Por ejemplo, consideremos una cantidad  $a$  que crece exponencialmente a una tasa  $b$  en una unidad de tiempo  $t$ . Que podemos escribir mediante la siguiente ecuación:

$$a(b, t) = a(1 + b)^t. \quad (1.5)$$

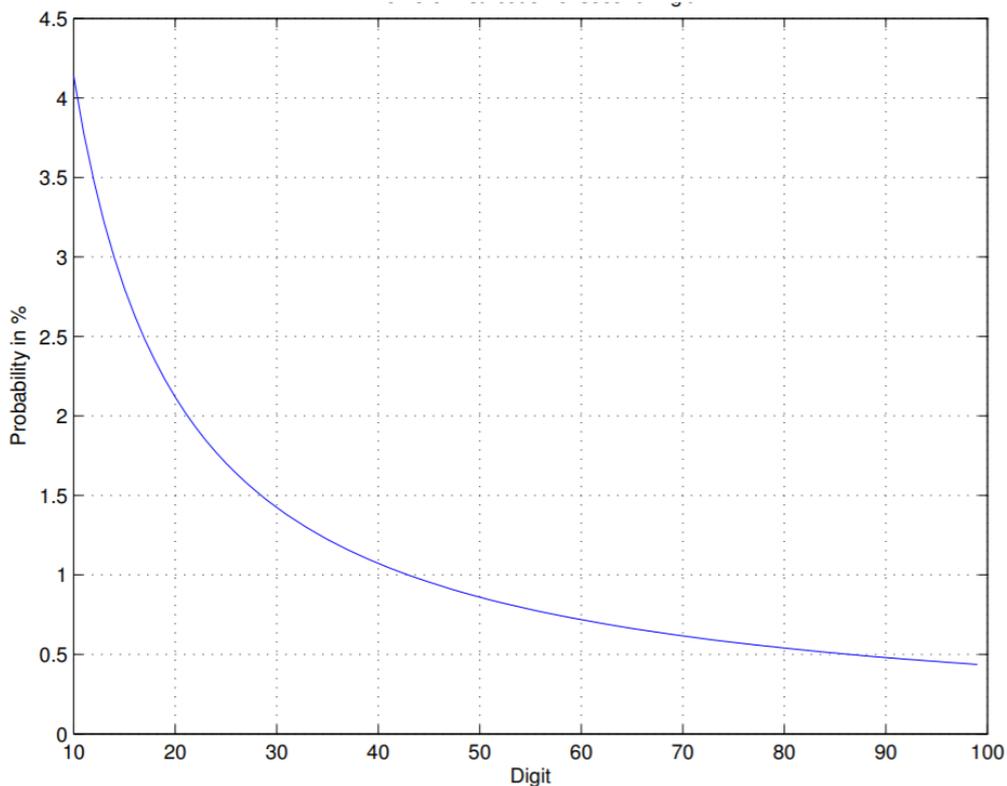


Figura 1.7: Probabilidades porcentuales de la aparición del segundo dígito en sistema decimal.

La Ec. 1.5 tiene la forma de una ley de potencias, que puede modelar muchos fenómenos de la naturaleza. Por ejemplo, físicos (La ley de Stefan-Boltzmann), biológicos (La ley de Kleiber que relaciona el metabolismo de un animal con su tamaño), geográficos, sociológicos, económicos, lingüísticos, entre otros.

Pues bien, como se ve en la gráfica, la variación de la frecuencia del primer dígito en función del tiempo muestra que los dígitos más pequeños son más frecuentes en un número mayor en el tiempo. Supongamos que  $a$  es igual a 10,000 y  $b$  representa el 4%. Entonces, se necesitan 18 unidades de tiempo para que el dígito 2 supere al 1 en la frecuencia de aparición del primer dígito. Luego, 11 unidades temporales para que el dígito 3 supere al dígito 2; si continuamos observamos que se necesitan tres unidades para que el dígito 9 supere 8.

Existen otros modos de explicar el mecanismo, este modelo de potencias, considero, es fácil de comprender pues es acumulativo, monótono ascendente y lo encontramos útil para representar muchos fenómenos de interés académico y práctico. Como es el conteo de votos de una elección presidencial.

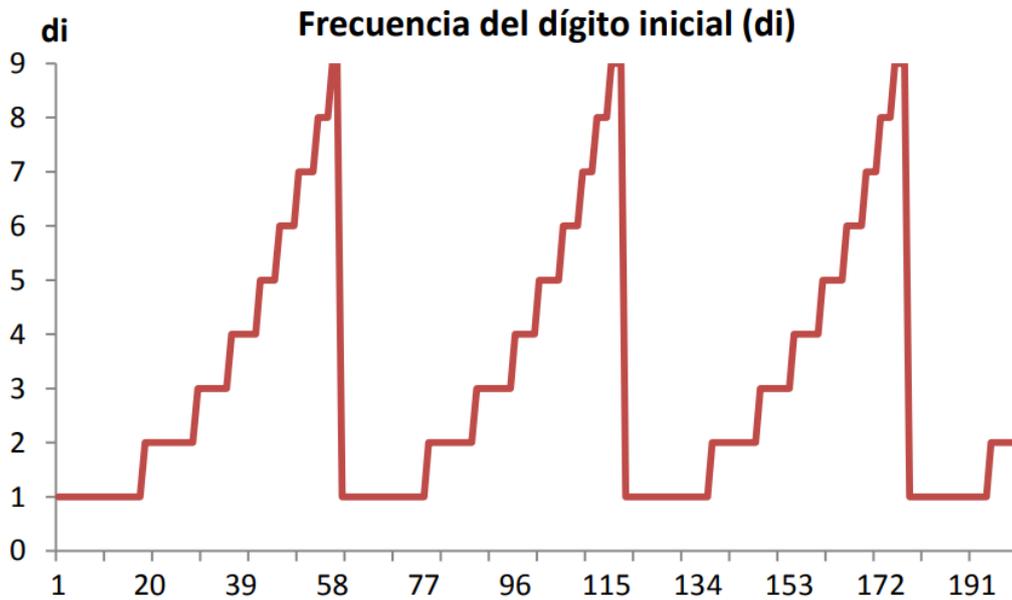


Figura 1.8: Probabilidades aparición de primer dígito en diferentes sistemas [20].

---

## 1.4. Una deducción matemática de la ley de Benford

Se ha probado que la ley de Benford sólo representa la distribución de probabilidad donde se observa invarianza ante el cambio de base o de escala, propiedades que hacen a la regla universalmente aplicable.

La invarianza de escala significa que si los primeros dígitos de serie  $x$  siguen la [21], entonces así pasará con los primeros dígitos de una variable reescalada  $kx$ , para cualquier valor de  $k$ . Ya que la distribución de Benford es la única con tal propiedad, si los primeros dígitos no siguen [21], entonces tampoco siguen el rescalamiento.

Se puede mostrar que un valor aleatorio  $x$  sigue una distribución uniforme o logarítmico (o equivalente) si la densidad de probabilidad  $P(x)$  es proporcional a  $1/x$ , entonces por integración simple de sus primeros dígitos seguirán la ley de Benford [21].

La ley de Benford se aplica a datos dimensionales, por lo que los valores numéricos de los datos dependen de las unidades. Si existe una distribución de probabilidad universal  $P(x)$ , entonces debe ser invariante ante un cambio de escala, de modo que:

$$P(kx) = f(k)P(x). \quad (1.6)$$

la función  $f(k)$  es parte de la solución que conecta con la ley de Benford, y es lo primero que deseamos encontrar en este desarrollo. Considerando la condición de normalización, obtenemos

$$\int P(x)dx = 1, \quad (1.7)$$

de manera que,

$$\int P(kx)dx = \frac{1}{k}. \quad (1.8)$$

Implicando que:

$$f(k) = \frac{1}{k}. \quad (1.9)$$

Ahora, desarrollaremos un poco más alrededor de esta idea para encontrar la función  $P(x)$ . Derivando con respecto a  $k$  y estableciendo que  $k = 1$ , obtenemos:

$$xP'(x) = -P(x), \quad (1.10)$$

---

la ecuación diferencial cuenta con la siguiente solución:

$$P(x) = \frac{1}{x}. \quad (1.11)$$

Si muchas potencias de 10 se encuentran entre los puntos de corte, entonces la probabilidad de que el primer dígito sea  $D$  se describe por distribución logarítmica:

$$P_D = \frac{\int_D^{D+1} P(x)dx}{\int_D^{10} P(1)dx} = \log_{10} \left( 1 + \frac{1}{D} \right), \quad (1.12)$$

con  $D$  que pertenece a  $1, 2, \dots, 9$ .

Para una lista de números siga una distribución de probabilidad en forma de ley de potencias  $N^{-1}$ , tendremos que la probabilidad del primer dígito significativo es independiente de la década y sigue la ley de Benford:

$$\begin{aligned} \int_{10^k(D)}^{10^k(D+1)} N^{-1} dB &= L_N(10^k(D+1)) - L_N(10^k D), \\ &= L_N \left( \frac{10^k(D+1)}{10^k(D)} \right), \\ &= L_N \left( 1 + \frac{1}{D} \right). \end{aligned}$$

Por último, normalizamos para escribir:

$$P(D) = \log_{10} \left( 1 + \frac{1}{D} \right). \quad (1.13)$$

Este resultado implica que las bases de datos que se distribuyen de acuerdo con una ley de potencia tienen la característica de ser invariantes de escala. Tanto su media como su desviación estándar (*i.e* divergencia) son parámetros insuficientes para caracterizar de una manera adecuada, pues se extienden en muchos ordenes de magnitud. De acuerdo con lo anterior, se establece que toda variable cuya distribución sea tipo ley de potencia tendrá propiedades similares a la ley de Benford. Es decir,

$$P(x, \lambda) \sim x^{-\lambda}, \quad (1.14)$$

donde  $\lambda > 0$  es un número positivo y real. Este resultado es el mismo del ejemplo mostrado en la Ec. 1.5.

## 1.5. Validez en otras bases numéricas

La Fig. 1.9 muestra la probabilidad de que aparezca el primer dígito en diferentes sistemas decimales. Por ejemplo, en el sistema binario ( $B = 2$ ) la probabilidad de que aparezca el 1 es de 100 %, pues no se suele escribir el cero antes del uno. Por ello siempre aparece el 1 como primer dígito. Al aumentar la base, también las opciones; por lo que disminuye la probabilidad de aparición del dígito 1.

Ahora, al estudiar bases mayores de 10, por ejemplo, el sistema hexadecimal (base 16), calcular la probabilidad de cada dígito sigue siendo simple, pero se debe cuidar de pensar que los números mayores que 9 (tal como del 10 al 15 en el caso hexadecimal) están representados por dos dígitos. Lo cual en definitiva es otro caso. En el sistema arábigo simplemente nos quedamos sin dígitos adicionales. Los informáticos utilizan «dígitos» alfanuméricos, por ejemplo, se asigna una letra A para el número 10, B para el 11, etcétera. Por lo tanto, cuando se solicita la probabilidad de que el primer dígito sea 12, entonces significa que se pide que el primer dígito sea C, y esto sigue siendo sólo un dígito, sin importar cuántos dígitos arábigos sean necesarios para representarlo. Así, la ley de Benford es válida para bases superiores a 10.

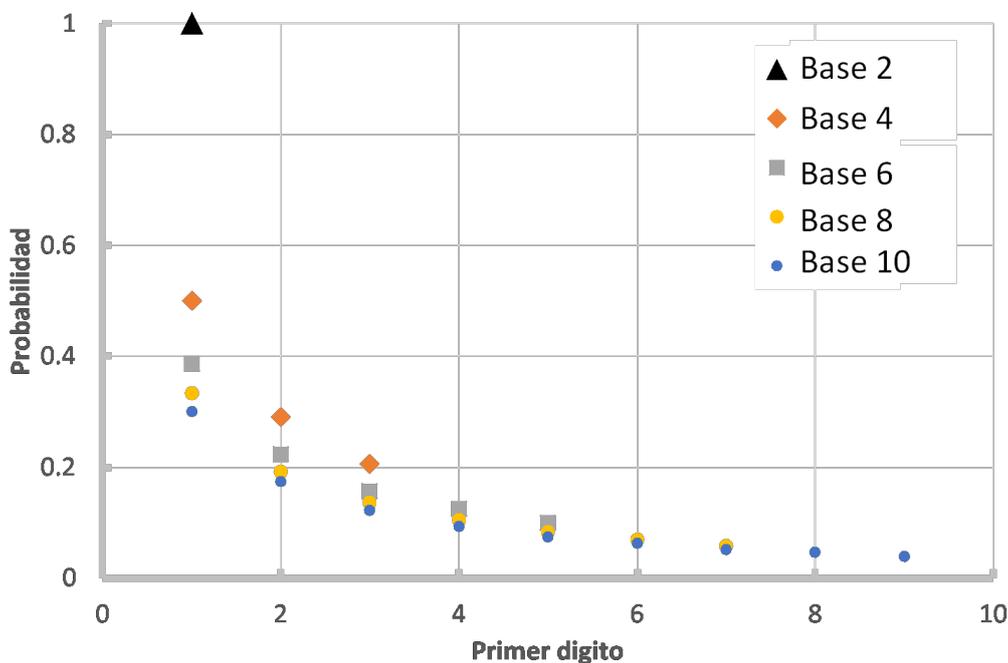


Figura 1.9: Probabilidades aparición de primer dígito en diferentes sistemas.

---

## 1.6. Limitaciones de la ley de Benford

En términos generales la ley Benford se cumple cuando los datos están distribuidos en varios órdenes de magnitud o distintas escalas. Pero aunque es una afirmación general y clara, también son términos muy vagos. Así que los expondremos con base a los supuestos principales de una distribución tipo Benford:

- Que sean sucesiones geométricas (requisito fundamental), es decir que no cuenten con máximo, ni un mínimo teórico; de otro modo, los números aparecerán con una marcada frecuencia. Por ejemplo, una serie de mediciones del largo de una mesa, tenderán a un valor central: la longitud de tal mueble. Por lo que los primeros dígitos estarán constreñidos a la medición (sin importar el número de mediciones) y sin ajustarse a la ley de Benford.
- Datos que contengan cuatro o más dígitos. Le da robustez a la colección de la base de datos. De otro modo el conjunto estudiado es inadecuado para observar la tendencia según Benford. Entonces, además de requerir una cantidad amplia de datos, cada dato debe ser consecuencia de un cálculo, medición o conteo amplio que supere cuatro cifras significativas.
- Datos que contiene valores similares para fenómenos similares. Si bien es posible combinemos bases de datos para obtener una tendencia similar a la ley de Benford, como se ilustra en la Fig. 1.6, lo mejor es mantener homogéneas o fieles las fuentes para evitar complicar la interpretación.
- El conjunto de datos no debe componerse de números asignados. El arquetipo de ejemplo son los números de teléfono, por zonas y regiones se asigna cierto número de identificación, que luego puede implicar otros código y finalmente un conteo. Así, si consideramos los números telefónicos dentro de un país y consideramos su clave, los primeros dígitos serán los mismos dentro de la base de datos. Es decir, tal distribución de primeros dígitos es uniforme.

De este modo, distribuciones como al lotería o la estatura de una población, o los números telefónicos no siguen la ley de Benford. Las estaturas no se extienden sobre muchos órdenes de magnitud sino que se acumulan en una escala específica, por ejemplo entre 140 y 190 cm. Esta clase de distribuciones se describen mejor por medio de la distribución Gaussiana.

Otro ejemplo, son los precios de productos obtenidos de diferentes comercios. Además de estar ceñidos a un par de decimales, oscilan alrededor de un

---

valor central. Cuando los números son usados como etiquetas suelen no seguir la ley de Benford, pero cuando son para medir si la siguen. En cambio, una variable que se extiende en varios órdenes de magnitud como la población de los países tienen un mejor ajuste con la distribución de Benford.

### 1.6.1. Otras limitaciones de la ley de Benford

Hill (1995) presenta una explicación convincente de por qué es muy común observar la ley de Benford en tablas numéricas generadas con fenómenos diversos. De acuerdo con Hill, la distribución logarítmica se establece cuando el conjunto de números proviene de un proceso aleatorio generado con una «mezcla estadística». Es decir, una tabla de números sigue la ley de Benford cuando éstos son producidos a partir de procesos aleatorios de un conjunto pequeño de distribuciones diferentes que, a su vez, son elegidos aleatoriamente. A manera de ejemplo, este escenario prevalece en una tabla recopilada con números que aparecen en una revista o periódico, ya que estos números provienen de diferentes fuentes de datos: estadísticas deportivas, información financiera, fechas de eventos, precios de productos enunciados.

Theodore P. Hill y Arno Berger argumentan que los datos deben mostrar ciertas características: independencia, invarianza de escala, invarianza de base, y contar con varios órdenes de magnitud. Sin embargo, es importante recordar, de nuevo, que cuando los datos provienen de diferentes fuentes, no homogéneas, al conjugarse pueden seguir la ley de Benford.

## 1.7. Conclusiones del capítulo

Newcomb fue el primer científico que descubrió la ley del primer dígito. Sin embargo el físico Franck Albert Benford le ayudó a darle más formalismo con una deducción logarítmica. Posteriormente otros matemáticos como Hill y Pinkman contribuyeron a generar una demostración más formal. No fue hasta la década de los 90s cuando Mark Nigrini usó dicha ley como aplicación para poder detectar fraudes Fiscales lo que contribuyó a que dicha ley logarítmica alcanzará su fama.

Si bien la ley de Benford se estudia para el primer dígito, se puede observar que se puede generalizar matemáticamente al segundo dígito y hasta llegar a una densidad conjunta. Sin embargo, al incrementar la posición del dígito la distribución tiende a ser uniforme.

Existen explicaciones empíricas y deducciones matemáticas para comprender mejor dicha ley. Con todo, a pesar de que la ley de Benford puede expresarse en varias bases numéricas siguiendo el mismo comportamiento

---

logarítmico su uso tiene limitaciones prácticas. En el siguiente capítulo expondremos de algunas de las principales aplicaciones generales de la ley de Benford.

## Capítulo 2

# Breviario de aplicaciones de la Ley de Benford

En este capítulo se resumen algunas de las aplicaciones (que considero) más relevantes y populares de la Ley de Benford. Para el estudio de la historia de la ciencia o de la epistemología la ley de Benford debe ser un caso interesante, pues primero se encontraron muchos ejemplos empíricos donde se cumple y mucho tiempo después se desarrolló la teoría que le sustenta matemáticamente.

### 2.1. Área físico-matemática

Si la ley de Benford en verdad se trata de un fenómeno intrínseco en la naturaleza, entonces (con mayor certeza) puede ser utilizado para detectar señales anómalas en los datos estadísticos electorales.

Como mencionamos, Benford se dio cuenta que muchas bases de datos siguen dicha ley logarítmica, algunos de estos datos provienen de mediciones de la naturaleza. La tabla 2.1 muestra 15 conjuntos de observaciones modernas de campos diversos, como son: matemáticas, física, química, astronomía, geofísica e ingeniería; todas ellas concuerdan de modo cercano a la ley de Benford [12]. (Si la ley de Benford en lugar de ser una ley estadística, en realidad fuese una ley física; entonces, se podría utilizar para comprobar el realismo de modelos matemáticos aplicados en procesos naturales, como sugieren algunas investigaciones) [22].

Para complementar la información de la tabla 2.1, elaboramos la Fig. 2.1 que muestra los datos que peor se ajustan linealmente a la ley de Benford (*i.e.* gravedad terrestre) y los que mejor se ajustan (*i.e.* series geométricas). Encontramos el peor coeficiente de determinación de 0.958 y equivale a la de

la masa de exoplanetas, pero tal valor de determinación considerada como aceptable para el área de la física. Sin embargo, al realizar ajustes lineales emparejando sólo nueve datos es inadecuado, pues carece de una buena representación estadística. En secciones más adelante mostramos que se pueden utilizar este mismo criterio considerando los dos primeros dígitos, lo que implica que se pueden analizar 100 datos.

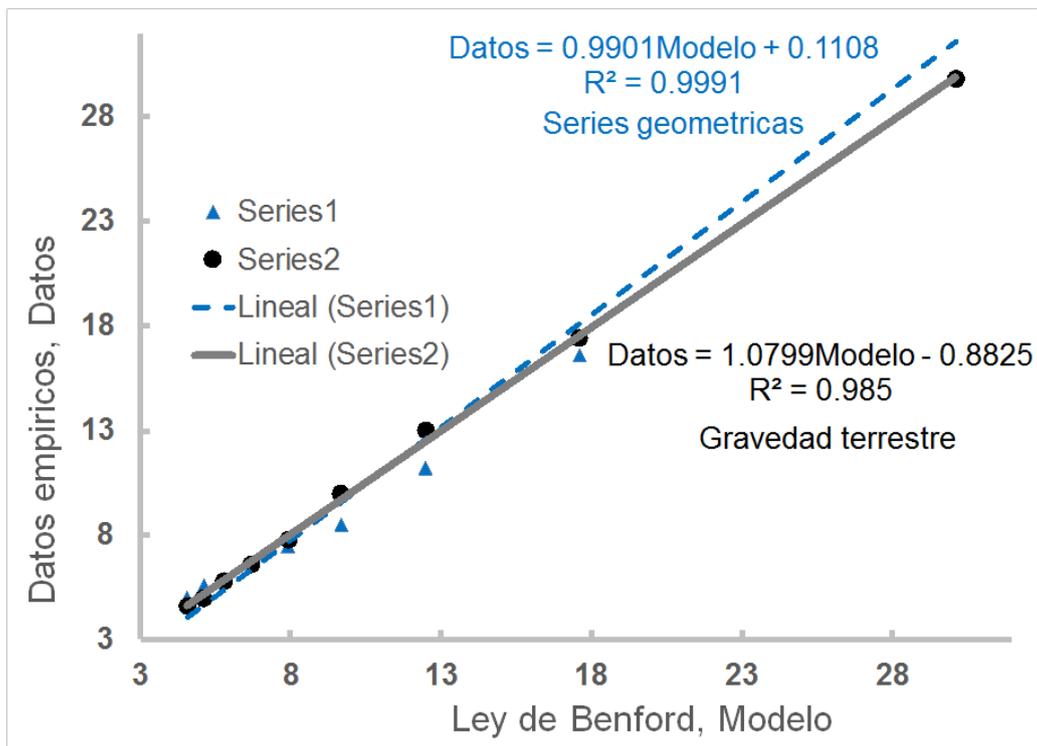


Figura 2.1: Gráficas que utilizan datos de la tabla 2.1, los datos que peor se ajustan a la ley de Benford: gravedad terrestre, contrastado con los que el que mejor la siguen: las secuencias geométricas.

En adición, parece que existen conjuntos de datos que se ajustan mejor que otros a la ley de Benford. En física existen tres distribuciones estadísticas ampliamente utilizadas, que son:

1. La distribución de Boltzmann-Gibbs, que es la probabilidad medida

Tabla 2.1: Distribución porcentual de los primeros dígitos de datos físicos y matemáticos, de acuerdo con datos de Sambridge [12].

|  | Frecuencia de los primeros dígitos |      |       |      |      |      |      |      |      | Número | Intervalo |
|--|------------------------------------|------|-------|------|------|------|------|------|------|--------|-----------|
|  | 1                                  | 2    | 3     | 4    | 5    | 6    | 7    | 8    | 9    |        |           |
| Ley de Benford                                       | 30.1                               | 17.6 | 12.49 | 9.69 | 7.92 | 6.69 | 5.80 | 5.12 | 4.58 | -      | -         |
| Campo magnético                                      | 28.9                               | 17.7 | 13.3  | 9.4  | 8.1  | 6.9  | 6.1  | 5.1  | 4.5  | 36512  | $10^{10}$ |
| Inversiones geomagnéticas                            | 32.3                               | 19.4 | 13.9  | 11.8 | 5.3  | 4.3  | 3.2  | 5.4  | 4.3  | 93     | $10^3$    |
| Velocidad de ondas sísmicas abajo del Pacífico-SO    | 30.0                               | 17.6 | 13.3  | 9.8  | 7.9  | 6.4  | 5.6  | 4.89 | 4.47 | 423776 | $10^6$    |
| Gravedad terrestre                                   | 33.0                               | 16.6 | 11.2  | 8.5  | 7.5  | 6.7  | 5.94 | 5.57 | 5.03 | 25917  | $10^9$    |
| Masa de exoplanetas                                  | 33.9                               | 15.4 | 10.7  | 9.2  | 6.23 | 9.47 | 5.98 | 4.48 | 4.48 | 401    | $10^5$    |
| Frecuencia de rotación de pulsares                   | 33.9                               | 20.7 | 12.7  | 7.6  | 5.3  | 5.0  | 4.94 | 4.67 | 4.88 | 1861   | $10^4$    |
| Flujo de rayos- $\gamma$ (telescopio espacial Fermi) | 30.3                               | 17.9 | 13.0  | 9.9  | 7.6  | 6.96 | 5.23 | 5.23 | 2.72 | 1451   | $10^5$    |
| Profundidad de terremotos                            | 31.6                               | 16.9 | 14.0  | 8.69 | 6.98 | 7.42 | 5.27 | 4.58 | 4.36 | 248915 | $10^2$    |
| Sismograma S-A                                       | 28.4                               | 15.7 | 12.5  | 9.6  | 8.97 | 7.37 | 6.52 | 6.04 | 4.93 | 24000  | $10^5$    |
| Emisiones invernadero por país                       | 29.9                               | 17.9 | 11.4  | 7.6  | 9.2  | 8.15 | 5.97 | 4.89 | 4.89 | 184    | $10^4$    |
| Anomalías de temperaturas globales entre 1880–2008   | 27.7                               | 19.4 | 12.7  | 12.1 | 8.9  | 5.4  | 6.61 | 4.32 | 2.81 | 1527   | $10^2$    |
| Constantes físicas fundamentales                     | 34.0                               | 18.4 | 9.2   | 8.28 | 8.58 | 7.36 | 3.37 | 5.21 | 5.52 | 326    | $10^4$    |
| Casos de enfermedades infecciosas (global)           | 33.7                               | 16.7 | 13.2  | 10.7 | 7.3  | 5.4  | 4.56 | 5.07 | 3.34 | 987    | $10^6$    |
| Series geométricas                                   | 29.8                               | 17.4 | 13.0  | 10.0 | 7.8  | 6.6  | 5.8  | 5.0  | 4.6  | 1000   | $10^{21}$ |
| Sucesión de Fibonacci                                | 30.0                               | 17.7 | 12.5  | 9.6  | 8.0  | 6.7  | 5.7  | 5.3  | 4.5  | 1000   | $10^{14}$ |
| Combinados   | 30.9                               | 17.4 | 13.2  | 9.0  | 7.6  | 6.4  | 5.7  | 4.8  | 5.0  | 10000  | $10^{33}$ |

---

usada para describir la distribución de estados en un sistema.

2. La distribución de Fermi-Dirac, que es una medida de las energía de partículas independientes que obedecen el principio de exclusión de Pauli *i.e.* fermiones.
3. La distribución de Bose-Einstein, una medida de la energía de partículas independientes que no obedece al principio de exclusión de Pauli, *i.e.* bosones.

Con respecto a la temperatura del sistema, tanto la distribución de Boltzmann y la de Fermi fluctúan de un modo periódico alrededor siguen la distribución de Benford. En contraste, la distribución de Bose-Einstein sigue la tendencia de la ley Benford independientemente de la temperatura. De hecho, las distribuciones logarítmicas son una característica de la física estadística, por lo que varios autores sugieren que podría significar algo más profundo para la estructura de la materia. [13], [14], [23].

Por ejemplo, un estudio estadístico de leptones- $\tau$  evidencia un discrepancia marginal a la ley de Benford, donde el dígito 9 presenta la mayor diferencia hasta en  $4\sigma$  [15].

En otro artículo académico [24], se menciona que el análisis estadístico de varios espectros de absorbencia infrarroja de varios polímeros también es congruente a la ley de Benford, pero es sensible cuando los datos de absorbencia son transformados a transmitancia óptica (la cual es una función acotada y logarítmica).

Con todo, un estudio más profundo sobre las relaciones entre datos físico-matemáticos y la ley de Benford (además de la teoría que los ligue) se encuentra fuera de los objetivos de esta tesis; pero la invitación esta abierta para futuros trabajos de investigación.

### 2.1.1. Detección de imágenes manipuladas

El gradual avance de las productos digitales presenta, entre otros fenómenos, la indeseable manipulación de las imágenes. Para confiar la autenticidad de una imagen se pueden emplear diversas herramientas y protocolos. Varios autores han propuesto utilizar la ley del primer dígito como un complemento analítico para validar imágenes digitales. Modelando la distribución de probabilidad del primer dígito y contrastando se ha propuesto su sensibilidad a la magnitud del gradiente de tonos, en especial al considerar código piramidal laplaciano y a la entropía de información [25], de modo que se puede comprobar la simple o doble compresión en imágenes de tipo JPEG es indiferente a la clonación de elementos, pero sensible a la incorporación de efectos ópticos

---

(como brillos); de modo similar, evidencia cambios en marcas de agua «semi-frágiles» [25]. Así, puede ser una herramienta para esteganografía, es decir, para detectar mensajes ocultos en imágenes [26]. Las aplicaciones forenses de la imagen digital son matemáticamente sofisticadas y de cambio constante en periodos de meses; pero para nuestros propósitos sólo ilustran como un sector tecnológico utiliza la ley de Benford.

### 2.1.2. Resolución de preguntas de opción múltiple

Algunas de las aplicaciones a la ley del primer dígito pueden ser curiosas. Por ejemplo, un estudiante con escaso conocimiento sobre física, pero actitud emprendedora, puede adoptar tal perspectiva para superar un examen de opción múltiple donde las respuesta y distractores son cantidades físicas, sin importar como se revuelvan las opciones. Es decir, después de un largo examen de preguntas de opción múltiple, cuando ya se carece de tiempo para verificar la respuesta, lo mejor puede ser escoger la opción que cuente con el primer dígito más pequeño, teniendo una mayor ventaja los que comiencen con 1, 2 o 3. Las pruebas con bancos de preguntas muestran que se tendría una ligera ventaja, de modo que debe evitarse como estrategia general en exámenes con pocas preguntas [16].

## 2.2. Área económica

En diversas actividades económicas encontramos aplicaciones de la ley de Benford, principalmente como un indicador de falsificación o alteración maliciosa de los datos.

La ley de Benford puede ser utilizada como una herramienta para detectar la manipulación indebida de los precios y en consecuencia del equilibrio en la curva de oferta-demanda de los insumos o bienes. Un caso relevante surgió en Europa en el año 2002, en esa época comenzaba a surgir el euro como la moneda corriente en la Unión Europea, de modo que varios vendedores minoristas comenzaron redondear los precios de los productos y bienes con el objetivo de convencer psicológicamente al consumidor de comprar un producto más barato [27]. El alza de precio de algunos productos puede ser considerada un delito en economías proteccionistas; pero en una perspectiva ideal de libre mercado este fenómeno sería aceptado.

Hal Ronald Varian, estadounidense especializado en microeconomía, argumenta que una de las aplicaciones de la Ley de Benford sería en las ciencias sociales. El autor argumenta que se puede usar esta ley para detectar posibles fraudes en datos socioeconómicos que presentan el gobierno, empresas

Tabla 2.2: Datos de primer-dígito aplicados a datos macroeconómicos [29].

| Dígito | Benford     | Home<br>Curr<br>Nominal | Home<br>Curr<br>Real | \$<br>(PPI)<br>Nominal | \$<br>(ExRat)<br>Nominal | \$<br>(PPI)<br>Real | \$<br>(ExRat)<br>Real |
|--------|-------------|-------------------------|----------------------|------------------------|--------------------------|---------------------|-----------------------|
| 1      | 0.3010      | 0.2737                  | 0.2856               | 0.3212                 | 0.3133                   | 0.3236              | 0.3138                |
| 2      | 0.1761      | 0.1756                  | 0.1582               | 0.1646                 | 0.1646                   | 0.1551              | 0.1867                |
| 3      | 0.1249      | 0.1242                  | 0.1282               | 0.0981                 | 0.1155                   | 0.1052              | 0.1131                |
| 4      | 0.0969      | 0.1084                  | 0.1052               | 0.0926                 | 0.1028                   | 0.0973              | 0.1028                |
| 5      | 0.0792      | 0.0910                  | 0.0823               | 0.0791                 | 0.0862                   | 0.0767              | 0.0736                |
| 6      | 0.0669      | 0.0704                  | 0.0815               | 0.0759                 | 0.0815                   | 0.0688              | 0.0665                |
| 7      | 0.0580      | 0.0609                  | 0.0625               | 0.0570                 | 0.0388                   | 0.0601              | 0.0570                |
| 8      | 0.0512      | 0.0554                  | 0.0530               | 0.0585                 | 0.0522                   | 0.0625              | 0.0427                |
| 9      | 0.0458      | 0.0403                  | 0.0435               | 0.0530                 | 0.0451                   | 0.0506              | 0.0396                |
|        | $\chi^2(8)$ | 8.7389                  | 9.1085               | 14.5519                | 15.8290                  | 13.3434             | 7.2416                |
|        | índice-p    | 0.3648                  | 0.3332               | 0.0685                 | 0.0449                   | 0.1006              | 0.5108                |

privadas y público en general. Pues, según Ronald, si la gente inventa los números estos no seguirán la ley de Benford, al parecer estamos propensos psicológicamente a tratar de mostrar un patrón uniforme [27], [28].

En cuestiones macroeconómicas, la ley Benford puede usarse para validar información de un país o un conjunto de países. Es decir, índices como el producto interno bruto, la tasa de desempleo, el déficit público, la balanza de cuentas corriente, entre otros pueden ser históricamente y geográficamente analizados desde la perspectiva de la distribución de Benford. La tabla 2.2 muestra dichos comportamientos económicos utilizando datos de 2 países de la Organización para la Cooperación y el Desarrollo Económicos OCDE, los valores de  $\chi^2$  e índice- $p$  muestran que algunos indicadores económicos se ajustan de modo aceptable<sup>1</sup> [29].

<sup>1</sup>Aunque de nuevo, sólo se analiza el primer dígito de la ley de Benford.

---

### 2.2.1. Aplicaciones en auditorías

El trabajo de la auditoría financiera consiste en revisar los estados financieros de las empresas. Ver su balance entre activos y pasivos, así como implementar estrategias para mantener las finanzas sanas; y por tanto, que la empresa crezca y se diversifique. Lo anterior requiere de un proceso ético ya que utilizando una adecuada información financiera es posible al inversionista decidir del mejor modo. Por ello, el trabajo del auditor es revisar que las cuentas de la empresa estén correctamente establecidas. La ley de Benford es útil como un criterio para que la auditoría se fundamente más; con la intención de prevenir sospecha y fraudes en los estados financieros.

El Instituto Americano de Contadores Públicos (por sus siglas en inglés: AICPA) ha considerado proponer una profesión que se dedique a la búsqueda de herramientas analíticas y métodos de auditoría para detectar el fraude. Específicamente, el SAS No. 99 (párrafo 28) y al igual SAS No. 56 al requerir a los auditores que empleen procedimientos analíticos durante la fase de planificación de la auditoría con el objetivo de identificar la existencia de transacciones, eventos y tendencias atípicas [30], [31]. Al mismo tiempo, SAS Núm. 99 advierte que confiar en procedimientos analíticos, que tradicionalmente se realizan en datos altamente agregados, puede proporcionar sólo indicaciones generales de fraude.

El propósito de tal documento es ayudar a los auditores en el uso más efectivo de la información usando el análisis digital basado en la ley de Benford. Cuando se usa correctamente, puede ayudar a los auditores a identificar cuentas específicas en las que podría residir el fraude. Puesto que desde hace tiempo, los auditores han aplicado diversas formas de análisis digital como la verificación de pagos duplicados o la búsqueda de números de cheques o facturas faltantes como métodos de procedimientos analíticos.

La ley de Benford aplicada a la auditoría es simplemente otra forma de análisis digital que consiste en la observación de una cuenta completa para determinar si los números recaen en la distribución esperada, es decir si se ajustan a una distribución teórica general o ley de Benford en particular.

Hal Varian (1972), economista de la Universidad de California, sugiere que la ley de Benford puede usarse como una prueba de la honestidad o validez de datos científicos en un contexto de ciencias sociales. Sin embargo, su visión no fue retomada por los contadores hasta finales de los años ochenta. Carslaw (1988) encontró que las ganancias de las empresas de Nueva Zelanda no estaban ajustadas a la distribución esperada. Más bien, los números contenían más ceros en la posición de segundo dígito de lo esperado y menos nueves, lo que implica que cuando una empresa tenía ganancias como \$ 1,900,000, redondeaban a \$ 2,000,000. Caslaw usó la distribución de Benford y la nombró

---

como «Prueba de Feller». Thomas, en 1989, descubrió un patrón similar en las ganancias de las empresas de EE.UU.

Nigrini parece ser el primer investigador en aplicar ampliamente la ley de Benford a los números contables con el objetivo de detectar el fraude. Según un artículo publicado en *Canadian Business* (1995), Nigrini se interesó por primera vez en el trabajo sobre manipulación de ganancias por parte de Carslaw. Thomas luego se encontró con el trabajo de Benford y combinó las dos ideas para su disertación. Usó el análisis digital para ayudar a identificar a los evasores de impuestos. Más recientemente, se han publicado documentos que detallan aplicaciones prácticas del análisis digital, como descripciones de cómo un auditor realiza pruebas en conjuntos de números contables, cómo un auditor usa programas de computación de análisis digital y estudios de casos para el entrenamiento de estudiantes.

## 2.2.2. Evasión de impuestos

Por otro lado la ley de Benford puede de ser utilizada para detectar el comportamiento de la conducta humana para facilitar la detección de evasión de impuestos. Pero para ello se pondrá en definición algunos concepto para entrar en contexto. La evasión no planificada (ENP) se define como una manipulación flagrante por el papel de impuestos de los artículos de línea en el momento de la presentación. La evasión planificada (EP) es el resultado de acciones planificadas para ocultar una pista de auditoría.

Christian y Gupta (1993) analizaron datos de impuestos para examinar la evasión secundaria. La evasión secundaria se deduce de sus datos y ocurre cuando los contribuyentes reducen el ingreso impositivo desde arriba del nivel de ingreso de la tabla de impuestos por debajo de una tabla de impuestos. Estos investigadores usaron la Ley de Benford para justificar la suposición de que los dos dígitos finales de la renta imponible deberían ser uniforme sobre el rango de [00, 99]. Su análisis mostró que un porcentaje de usuarios de tablas de impuestos superior al esperado se posicionó en dólares superiores de los soportes de la tabla.

En julio de 1995, el Ministerio de Finanzas de los Países Bajos analizó las frecuencias digitales de los datos seleccionados de la declaración de impuestos. El objetivo del análisis fue detectar la falta de cumplimiento, evaluar la viabilidad de incorporar pruebas digitales seleccionadas en un modelo de selección de auditoría, y evaluar si los informes de terceros contribuyen al cumplimiento. Se analizó una muestra de 30,000 declaraciones de impuestos para el año fiscal de 1992. El análisis incluyó una tabulación, con gráficos, de los datos recibidos de interés por contribuyente, de acuerdo con los retornos de terceros de los bancos holandeses a las autoridades tributarias. Las fre-

---

cuencias de primer y segundo dígito mostraron una conformidad casi perfecta con la Ley de Benford.

La ley Benford es utilizada como una distribución esperada para los dígitos en datos tabulados. La suposición son los dígitos de los datos que se informaron con sinceridad o que están sujetos a EP. Deben conformarse a las frecuencias digitales esperadas. Se desarrolla un modelo de Factor de Distorsión que cuantifica la extensión de ENP.

### 2.3. Área electoral

La generación de esta ley en datos socio-económicos y de otra índole depende de la presencia de un doble proceso aleatorio: eventos de una distribución y distribuciones de probabilidades elegidas de un conjunto de datos reducido. Walter Mebane ha empleado esta metodología para detectar fraudes electorales, por lo general estudiando el primero o segundo dígito inicial en la cuenta de votos recibidos por cada candidato, ya sea en la casilla o en las secciones electorales [2].

Básicamente el cuestionamiento obedece a la dificultad de saber, efectivamente, si las decisiones de voto producen una distribución logarítmica en los primeros dígitos de las cuentas de las diferentes casillas o distritos. En especial, si pensamos que los procesos electorales son caracterizados como sistemas adaptables complejos.

En consecuencia, para que la ley de Benford funcione en datos electorales la decisión de votar debe contar con explicaciones de fuentes en términos de colecciones de funciones de distribución. Por ejemplo, un voto emitido en



Figura 2.2: Retrato fotográfico de Walter R. Mebane Jr., registrado aproximadamente en el año 2018. Mebane ha concentrado sus investigaciones en detectar fraude electoral por medios estadísticos.

---

determinada casilla debe provenir, en ciertas ocasiones, de un votante duro, en otras, de la decisión de un individuo con un compromiso partidista débil; mientras que otros más serían emitidos por personas independientes atraídos por lo menos por un rasgo del plan de campaña o el candidato o bien es otra la razón. En cualquier caso, la decisión de votar por un candidato está determinada por la distribución de probabilidad específica. De esta forma, el tipo de votante y la decisión que adopta son el resultado de un doble proceso de aleatorización.

Algunos autores han sugerido que las leyes de dígitos significativos no son relevantes para datos electorales. Por ejemplo, Taylor (2005) al igual que otros politólogos cuantitativos, plantea que la decisión de votar puede describirse como un esquema de lotería. De manera, que la probabilidad de un individuo de votar por un candidato se define en términos de un modelo multinomial y por ello existen diferentes distribuciones dependiendo de variables sociodemográficas, ideología y otros elementos que sirven para categorizar la población de votantes; por tanto, la cuenta total de votos es la suma de una serie de realizaciones de eventos aleatorios. Así, para el caso de las elecciones presidenciales de Venezuela en el 2018, el autor encuentra una discrepancia entre los datos y la ley de Benford, pero suponiendo que las elecciones fueron limpias si se ajustan a un modelo multinomial de exclusivamente variables que reflejan un proceso limpio; por lo que el estudio puede estar sesgado por la ausencia de variables que empañen el proceso.

### **2.3.1. Caso: Elecciones mexicanas-2006**

Las elecciones presidenciales realizadas el 6 de julio del 2006, han sido unas de las más reñidas en la historia de México. Tanto el proceso como los resultados de la elección fueron cuestionados mediante el alegato de un posible fraude electoral. Al día siguiente de la elección, el Dr. en Física Luis Mochán (investigador de la UNAM) destacó, numerosas inconsistencias en el comportamiento del reporte temporal de los datos por parte del Instituto Federal Electoral IFE. Al poco tiempo, el Dr. en Física Victor Romero Rochín y colegas de la UNAM, advirtieron de lo anómalo del comportamiento de los datos presentados por el IFE, entre las que destacaron la improbable falta de estabilización de los porcentajes acumulados de los partidos políticos que contaban con más votos PAN y PRD. El 4 de agosto de ese año, se realizó un seminario especial en el Instituto de Geofísica de la UNAM, congregando a 61 investigadores de diversas disciplinas pertenecientes a 25 instituciones del país, tras discutir los resultados de trece estudios realizados en México y uno en la Universidad de Cornell, EE.UU., con base en los datos publicados por el IFE, concluyeron una falta en el principio de certeza; por lo que consideraban

imprescindible la realización nuevamente de un conteo de todos los votos. Este grupo de investigadores publicaron el manifiesto titulado: «Científicos por un conteo de los votos que garantice la legalidad y la certeza de las elecciones del 2006».

Entre los trabajos presentados de tal reunión de trabajo el Dr. en Matemáticas Ricardo Mansilla empleó la distribución de Benford para estudiar 65,563 casillas. Para tal investigador los datos empíricos no siguen la ley de Benford, hay un rechazo estadístico que ponen los resultados electorales en entredicho. El Dr. Masilla hace hincapié que los dígitos 8 y 9 de la distribución del candidato Andrés Manuel López Obrador se ajustan adecuadamente a la ley de Benford; mientras que en los primeros dígitos hay poco acuerdo con tal ley. Además, el investigador argumenta que existen fluctuaciones en los estimadores muestrales debido a la finitud de las muestras haciendo pruebas de hipótesis y calculando la zona de rechazo. Tales reportes son públicos por Internet, al día 28/01/2020 son accesibles [32].

En un análisis mucho más riguroso, Mebane (2006a) propone la prueba de segundo dígito de ley de Benford para las elecciones presidenciales en México. El autor presenta lo adecuado que puede ser el ajuste de la distribución empírica para los 32 estados y para los datos globales. Con 95 % de confianza, el autor encuentra que la ley de Benford se rechaza estadísticamente en los datos globales y en muchos de los estados.

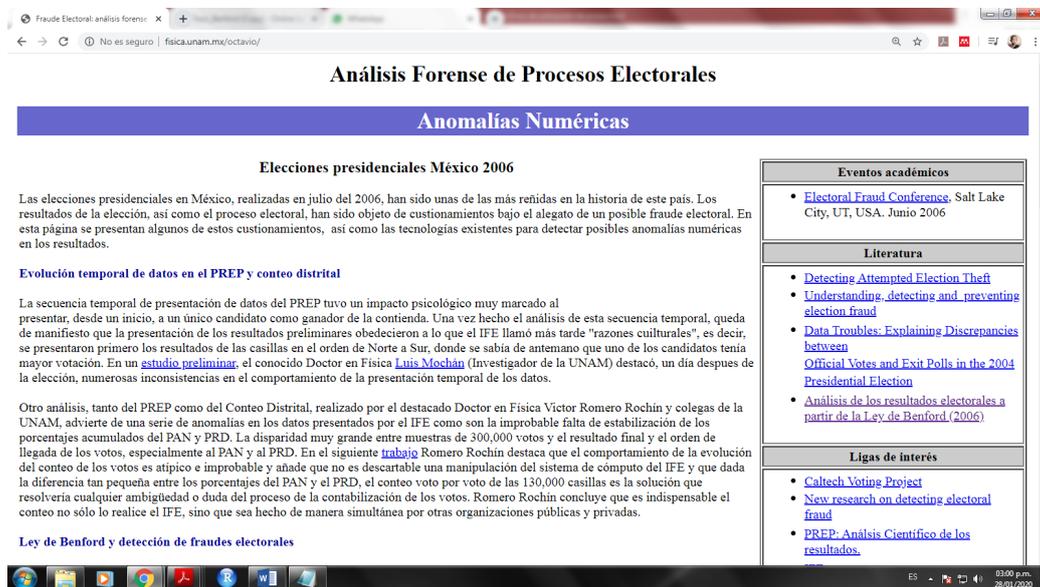


Figura 2.3: Captura de pantalla del sitio llamado «Análisis Forense de Procesos Electorales», último acceso 28/01/2020.

---

En otro artículo, Mebane (2007a) estudia los resultados de las elecciones de presidente, senadores y diputados de 2006 usando las cuentas por secciones. En su análisis encuentra muchas inconsistencias con el modelo de segundo dígito de la ley de Benford en los votos para los candidatos del PAN y PRD en municipios cuyo alcalde tiene la misma filiación política. Para este autor, dicho resultado ponen en evidencia que el partido dominante a nivel local operó en la fabricación de votos; no obstante, este patrón no se detectó en municipios gobernados por el PRI.

En un estudio más, Guitierras y Calderon (2006) presentan pruebas basadas en el segundo dígito de ley de Benford para resultados por sección analizados a nivel distrito. Los autores detectaron 47 distritos con severas anomalías.

Con todo, algunos autores polemizan sobre la utilización uso de ley de Benford para estudiar el fraude electoral. Mencionaremos un par de trabajos, con el único fin de mostrar que una auditoría de bases de datos por medio de un análisis estadísticos –como es la regla del primer digito– no es un proceso estándar del que se entienda en su alcance. Pliego (2007) argumenta que la violación de la ley de Benford es un mal indicador de fraude electoral. En su trabajo, el investigador analiza exclusivamente las casillas en las que el tribunal electoral decidió recontar los votos. Por tanto, Pliego plantea que el estudio del segundo dígito de ley de Benford puede ser considerada como una prueba potente sólo cuando la desviación observada en la frecuencia empírica respecto a la distribución teórica se incrementa conforme aumenta el número de irregularidades detectadas en el recuento. Es decir, la ley de Benford es un complemento para la carpeta de investigación a presentar ante la corte.

De acuerdo con la  $\chi^2$  calculada para 12 de los 15 distritos electorales involucrados (usando datos a nivel casilla y sección), es inobservable patrón alguno entre el valor de esta estadística y las modificaciones en los votos obtenidos. De aquí, que la ley de Benford sea descartada. Cabe mencionar que una limitación del enfoque de Pliego es que la desviación respecto a estudio de segundo dígito de ley de Benford puede deberse a la presencia de intimidación del voto y no a inconsistencias aritméticas. De acuerdo con Gonzalo Castañeda, la ley de Benford no es una prueba robusta para distinguir entre elecciones limpias de las que han sido manipuladas; por lo cual se debe utilizar con reservas.

En todo caso, es claro que las elecciones presidenciales mexicanas del 2006 mostraron a un gran conjunto de la población la relevancia del análisis estadístico de resultados electorales. Por ello, con el fin de evitar desinformación de los alcances de la técnica, a la vez para desarrollar la herramienta a un nivel de mejor explicación y mayor confiabilidad en sus conclusiones tratamos de aportar en este trabajo de tesis.

---

## 2.4. La estadística en el contexto forense

Los investigadores forenses, principalmente los criminólogos con conocimientos en análisis de datos, aprovechan la estadística para identificar patrones sociales relacionados con el crimen. Si bien puede ser polémico algunas conclusiones audaces realizadas por antropólogos sociales y economistas, pues derivan en políticas sociales radicales, también es cierto que ha encontrado utilidad para prevenir tragedias. Por ejemplo, al definir reglas de seguridad en trasportes aéreos, terrestres y marítimos. Por el momento, no abundaremos más en el tema, por ser amplio y por estar fuera de los objetivos de esta tesis.

En comparación con otros campos científicos, la ciencia forense es *sui generis*. La ciencia forense es inherente multidisciplinar, en especial por ser una ciencia solicitada y aplicada para resolver en el ámbito jurídico. Así en este

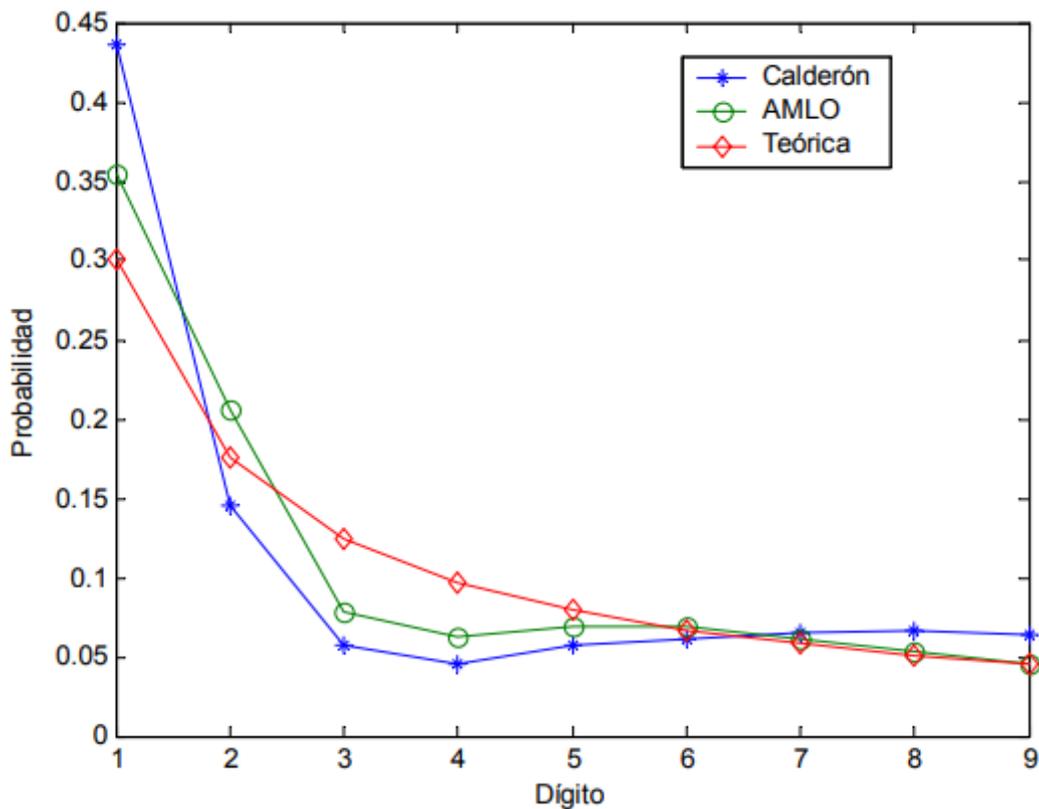


Figura 2.4: Ley de Benford contrastado con datos electorales de AMLO y Calderón, de acuerdo con la figura 1 presentada en el reporte de R. Mansilla (2006) [32].

---

contexto electoral, convergerían perfiles relacionados con las matemáticas, las leyes, psicología de masas, entre otros.

En un contexto de información parcial o limitada, la ciencia forense se interesa en indagar sobre las causas a partir de la observación de los efectos, con el fin de obtener las causas del pasado<sup>2</sup>. Más aún, la ciencia forense muestra dos rostros: uno académico y otro de operador. Por ejemplo, las investigaciones de los peritos están constreñidas a responder puntualmente las preguntas del ministerio público (quien tiene un perfil profesional de abogado); de modo que este operador carece de autoridad para ahondar en la investigación. En la academia, el científico cuenta con más libertad para indagar sobre su objeto de estudio.

Por otro lado, existen recomendaciones –de impacto internacional– para los jueces y abogados para reconocer la pertinencia de la información científica; como el llamado estándar Daubert [33], que consta de cuatro recomendaciones generales, que listo de modo resumido y laxo:

- **Falseable.** Que la teoría o la técnica se pueda reproducir o sea demostrable.
- **Publicidad.** Que la teoría o la técnica haya sido publicada o bien fuera sido revisada por miembros de la comunidad científica.
- **Tasa de error.** Que se conozca el margen de error en el cual se puede incurrir.
- **Aceptación de la comunidad científica.** Que la comunidad científica del área acepte la teoría o la técnica.

Todo este contexto afectar la indagatoria de un posible fraude electoral, por lo que debe ser considerada por quien utilice estadística como herramienta en tal estudio. Así, si se decide utilizar la ley de Benford, se debe contar con más que las credenciales técnicas en matemáticas, programación; se debe estar consiente del contexto forense y su responsabilidad. Por su parte, es insuficiente ser parte del quehacer forense, evidentemente se necesitan conocimientos técnicos y matemáticos para entender los alcances propuestas como la detección de fraudes por medio de estudio de índices estadísticos.

---

<sup>2</sup>Muchos otros campos estudian el pasado, pero sus escalas de tiempo son más lejanas: historia, paleontología, cosmología, entre otras.

---

## 2.5. Conclusiones del capítulo

La ley de Benford se cumple para diversas y múltiples bases de datos, lo que la hace atractiva para convertirse en un tipo de prueba de fiabilidad o detección de engaño en listados económicos y contables, como lo son los resultados electorales. Es claro que existe una amplia demanda para su consideración forense.

Sin embargo, el contexto forense presenta consecuencias que exigen que los analistas, abogados y jueces entiendan las limitaciones de este método. En el siguiente capítulo abundamos en el uso de pruebas estadísticas con el fin de encontrar malversación de datos electorales.

## Capítulo 3

# Análisis estadístico para identificar irregularidades electorales

En este capítulo se presenta una crítica más amplia sobre la aplicación de la estadística y probabilidad para identificar fraude electoral. Si en el capítulo anterior se expusieron las bases matemáticas y aplicaciones generales de la ley de Benford, aquí, como tópico central, se muestra su uso específicos para estudiar bases de datos de resultados electorales y sus alcances, si es que se desea presentar una demanda ante la corte. Si bien, en el capítulo anterior se trato brevemente el asunto, creo que se requiere profundizar la explicación pues esta tesis trata en particular el tema del estudio de bases electorales.

Estos párrafos son una síntesis y análisis sustentado principalmente en el reporte preliminar «Sobre la identificación de irregularidades electorales por medio de métodos estadísticos» redactado por la Comisión de Venecia para la *Comisión europea por la democracia vía la ley* del año 2018.

### 3.1. Malas prácticas electorales

Las controversias sobre la legitimidad e integridad de elecciones pueden desencadenar tumultos, revueltas e incluso la guerra civil. Así, el escepticismo electoral ha fomentado la investigación académica y el desarrollo legislativo en aras de mejorar métodos que legitimen la rectitud de las elecciones. En este trabajo, consideramos que cualquier fraude o error socava la integridad de las elecciones como libres, justas y competitivas. Por ello se debe evitar cualquier manipulación o pifia al respecto.

Existen diversas maneras en que los resultados de una elección pueden

---

ser manipulados o ser erróneos; en contraparte, también existe una gran cantidad de procedimientos forenses para identificar los diferentes tipos de irregularidades electorales. Entre tales procedimientos forenses, encontramos los métodos estadísticos para examinar los resultados electorales. La mayoría de estos métodos de análisis de base de datos buscan indicios de manipulación o alteración, de modo que es una perspectiva específica para abordar unos pocos tipos de irregularidades. Los patrones numéricos construidos por humanos tienden a diferir de aquellos que ocurren con «naturalidad» en conteos genuinos. De modo que se han desarrollado procedimientos estadísticos para revelar tales diferencias, pero son incapaces de distinguir con absoluta perfección de los resultados impolutos de aquellos que fueron construidos mediante la alteración de la base de datos original. La presente tesis considera este punto de vista de investigación. Otras perspectivas de estudio estadístico están fuera de los objetivos de este trabajo; tales como la participación ciudadana, los votos inválidos y la uniformidad de los votos en diferentes casillas electorales (lugares donde se emite el voto) geografía política, etcétera. Para ambos casos, encontrar variaciones en diferentes casillas es natural, por lo cual se debe evitar la implicación automática del delito o el error del proceso global.

La integridad electoral requiere una serie de condiciones y etapas, por lo que su pérdida puede ocurrir por muchas razones. El fraude electoral es más frecuentemente asociado con irregularidades el día de la elección, incluyendo la supresión o la intimidación de los votantes, rellenar urnas con la inserción de boletas extra<sup>1</sup>, alteración del contenido de las urnas, conteo deliberadamente equivocado de los votos, o bien un reporte erróneo de las cuentas. Sin embargo, la integridad de la elección también puede ser cuestionada en otras etapas del proceso electoral. Entre las prácticas fasinerosas más comunes encontramos:

- Restricción u obstaculización ilegal en el registro de candidatos y votantes.
- Restricción al derecho de congregación, reunión, durante campañas o en el uso de recursos públicos y medios para realizar campaña.
- Uso de la coerción o amenazas para afectar la participación de los ciudadanos en las elecciones
- Compra de votos.
- Sesgo partidista en las reglas electorales.

---

<sup>1</sup>Proceso conocido como «urnas embarazadas».

- 
- Alteración del relleno de la boleta o urna electoral.
  - Manipulación del recuento de elecciones o del informe del recuento.
  - Evitar que los funcionarios electos asuman el cargo.
  - Privación de los cuerpos electos de su poder de decisión.

Los agentes que comenten estos delitos pueden ser los administradores de una elección nacional o local. Pero también autoridades públicas u otros actores (*e.g.* proveedores de servicios, fabricantes de hardware, grupos paramilitares, colectivos a favor de un partido específico, opositores, hackers, gobiernos extranjeros, por mencionar algunos) también pueden asumir un rol en la manipulación de la elección.

La adulteración puede ser el mayor impedimento para la integridad de las elecciones, pues puede «legitimizar» a un gobierno que no representa la preferencia de la mayoría de los votantes.

Después de la llamada Guerra Fría<sup>2</sup> tanto académicos como políticos han propuesto diferentes medidas para salvaguardar el voto ciudadano. La más común es realizar una misión de observación del proceso electoral, o bien denunciar a través de medios de comunicación masivos y las formalización de acusaciones ante las cortes internacionales: en conjunto han permitido marcos de referencia sobre lo que se considera un proceso justo, más que una «jornada electoral limpia».

Algunos investigadores se han enfocado en los métodos estadísticos para identificar anomalías en los resultados electorales. Algunos apellidos prominentes de tales investigadores son: Myagkov, Mebane y Deckert. La técnica general para identificar irregularidades electorales por medios estadísticos es la selección de una base de datos fuente, formular una hipótesis sobre cómo los resultados de la elección deberían generarse ante la ausencia de fraude, procesar los datos y emitir un dictamen. Los datos en los que se basan estas pruebas suelen ser de dos clases:

- Fuentes de datos observables. Donde se incluye el recuento de votos a diferentes niveles geográficos –desde una casilla local hasta la tendencia de un país–, datos de votos válidos e inválidos, número de votos válidos, datos de participación electoral, coherencia con registros nacionales – como el censo, el registro nacional de nacimientos, defunciones, entre otros.

---

<sup>2</sup>Periodo histórico situado entre 1947 y 1985.

- 
- Fuentes de datos no-observables, que pueden contener información oculta para los analistas. Entre las que se encuentran: listas de observadores y agentes del partido por casilla, las restricciones (naturales o de otra índole) de acceso a los lugares de votación, entre otras.

Entonces, es posible utilizar hipótesis estadísticas sobre datos observables para estudiar el desempeño de una elección. Algunas hipótesis, sin embargo, pueden requerir información de fuentes de datos no-observables. Por ejemplo, si se realiza una elección entre dos candidatos A y B en sólo tres mesas de votación, de modo que dos mesas electorales muestran al candidato A como ganador con porcentajes de 60.5 % y 57.1 %, mientras que la mesa tercera presenta al candidato B como ganador con 80 %. Estos resultados podrían parecer sospechosos por la discrepancia; suponiendo que los votantes fueron asignados de manera uniforme a los colegios electorales, porque se esperaría preferencias de votación comparables en las tres casillas. Ahora bien, si los votantes están distribuidos siguiendo una regla entre los colegios electorales, como puede ser alfabéticamente, puede ser que varios miembros de una misma familia que votaron por B se registraron con la mesa tercera, proporcionando una explicación de la discrepancia. Suponiendo que los miembros de una misma familia tienden a presentar preferencias electorales similares. Para un análisis estadístico, la información de los votos inválidos e válidos para cada mesa electoral son datos observables, pero para convertirse en un comprobable mediante hipótesis estadística, también debe ser considerada la información no-observable sobre cómo se asignan los votantes a la mesa electoral.

Tal vez, el factor crucial de cualquier análisis de manipulación electoral es el conocimiento de los adversarios y sus capacidades. Un adversario podría ser un partido gobernante que busca permanecer en el poder; también puede ser un presidente de una mesa electoral, a quien le gustaría alterar el resultado; o bien podría ser un régimen extranjero que busca interrumpir al gobierno o instalar a sus propios dirigentes.

### **3.2. Alcances del análisis estadístico forense de elecciones**

Con el fin de identificar patrones atípicos entre los números, se han desarrollado pruebas estadísticas para analizar resultados electorales. Las irregularidades en los resultados electorales pueden ser la conclusión de manipulaciones en el proceso electoral o durante el recuento de votos. Los tres

---

ejemplos siguientes representan diferentes tipos de fraude que pueden detectarse con un procedimiento estadístico apropiado, diseñado para ser sensible a esa forma particular de mala práctica electoral.

La mayoría de la literatura académica da cuenta de pruebas centradas en el proceso de conteo de votos para detectar patrones considerados poco probables en conteos genuinos, pero que podrían ser un subproducto de los resultados de falsificación. Una segunda familia de pruebas busca correlaciones entre diferentes aspectos de los resultados de la votación. Este segundo tipo de pruebas puede mostrar que el intervalo de fraude es más amplio. Por ejemplo, podrían detectar la invalidación sistemática de las boletas durante los recuentos de votos en algunas áreas, patrones más grandes de relleno de las urnas por parte de los funcionarios electorales, o incluso la coerción local de los votantes para participar y votar por un partido específico. Por lo tanto, esta segunda familia podría detectar mala práctica electoral que ocurrió el día de las elecciones. El alcance va más allá del simple recuento. Sin embargo, estas pruebas se basan en suposiciones más fuertes sobre el mecanismo que genera resultados electorales limpios. A su vez, eso conlleva un mayor riesgo de falsas alarmas (falsos positivos, por decir de modo más técnico) por patrones que surgieron de procesos naturales, en lugar de fraude.

Debido a que la mala práctica puede alterar los resultados electorales en muchos puntos del proceso electoral, existe la necesidad de un espectro de pruebas estadísticas sensibles a los problemas en diferentes aspectos de la elección. Desafortunadamente, el uso de una multitud de pruebas también aumenta el riesgo de marcar por error una elección justa como fraudulenta, como resultado de la «multiplicidad estadística».

No obstante, existen algunos tipos de mala práctica electoral que no pueden detectarse mediante pruebas estadísticas y que sólo pueden identificarse mediante información cualitativa. Esto incluye condiciones injustas para las campañas electorales; irregularidades en el registro de candidatos o votantes; procedimientos nacionales sistemáticos de invalidación de boletas u otras actividades que afectan los resultados de la votación en todo el territorio y limitaciones en el ejercicio del mandato.

### 3.3. Consideraciones estadísticas

Decidir si los resultados electorales son incorrectos o si han sido manipulados, alterados o falsificados equivale a un problema binario de clasificación: etiquetando la elección como «mancillada» o «limpia». «Mancillada» podría significar que el resultado electoral es incorrecto. Es decir, que los ganadores anunciados, en realidad, son perdedores, o simplemente que los números re-

---

portados son incorrectos, según el contexto. «Limpio» podría significar que los números reportados son correctos o que los ganadores anunciados en verdad son los triunfadores, a pesar de cualquier error o malversación que pueda haber ocurrido.

A saber, en todos los problemas de clasificación binaria pueden ocurrir dos tipos de errores: clasificar por error un elemento que es de la primera clase como perteneciente a la segunda clase, y viceversa. En estadística, la clasificación binaria a menudo se formaliza por medio de una prueba de hipótesis. Un contexto al problema de clasificación en el análisis forense electoral establecería que la hipótesis nula es que los resultados son limpios. En contraste, la hipótesis alternativa que sería lógicamente la negación de la hipótesis nula es decir que los resultados no son limpios. Un falso positivo (conocido como error estadístico alfa o de Tipo I) es etiquetar una elección limpia como «mancillada»; un falso negativo conocido como error de Tipo II, es etiquetar una elección contaminada como «limpia». Si hubiera un modelo teórico-estadístico confiable que describiera el proceso generador de resultados electorales en ausencia de error o manipulación, se podrían desarrollar métodos que cuenten con una probabilidad máxima de clasificar erróneamente una elección «limpia» como «mancillada». Pero tal marco teórico (pese a ser deseable) es inexistente.

Regresemos al planteamiento del problema de clasificación como una prueba de hipótesis. La hipótesis nula es que la elección es limpia; como alternativa, su negación es decir la elección no es limpia. Entonces, se podrían construir pruebas que presenten la probabilidad máxima que permita dictaminar que una elección considerada limpia, en realidad esta mancillada.

Por desgracia, las probabilidades que surgen en los métodos forenses numéricos no son correctas simplemente por consenso, como lo son las probabilidades en las auditorías de limitación de riesgos. Esto se debe a que las probabilidades no surgen porque el auditor seleccione una muestra aleatoria de forma controlada y deliberada, ni de ningún otro mecanismo conocido o dominado por el investigador. Más bien, las probabilidades calculadas derivan de suposiciones sobre la distribución de probabilidad de resultados de elecciones limpias: cómo los números «deberían» comportarse en ausencia de error, fraude o manipulación, y cómo ese comportamiento varía o no con otros factores.

Estos supuestos estadísticos no se basan en verdades fundamentalmente establecidas sobre referencias políticas: es inexistente suponer una «naturaleza» confiable de las elecciones. Más bien, los supuestos son hipótesis de trabajo de mayor o menor plausibilidad que conducen a predicciones que se pueden comparar con otros datos y otras elecciones.

La evidencia de que los reportes son estadísticamente inconsistentes con

---

esos supuestos estadísticos representa un misterio: ¿por qué este conjunto de resultados parece peculiar? La solución a esa pregunta no debe condicionarse a la irregularidad electoral; podría ser simplemente que esos supuestos más o menos plausibles no son (aproximadamente) verdaderos en esa elección. Por ejemplo, muchos métodos de detección de irregularidades implican suposiciones sobre la distribución de dígitos en los resultados electorales correctos. Una clase de métodos supone que, en ausencia de irregularidades, los dígitos terminales se distribuirían uniformemente y con independencia entre conjuntos de casillas.

Otros métodos hacen suposiciones sobre las preferencias políticas. Por ejemplo, que el voto compartido que recibe un candidato es independiente sistemáticamente del número de votos emitidos en el lugar de votación, o que las preferencias políticas son independientes de la tecnología que utiliza una jurisdicción para contar los votos. Demostrar que tales hipótesis son inconsistentes con los datos no implica que haya alguna irregularidad. Por el contrario, un adversario experto siempre puede fabricar resultados fraudulentos de tal manera que los números pasen cualquier conjunto de pruebas estadísticas. Hay indicios de que esto sucedió en un caso donde la media de los segundos dígitos se ajusta a las predicciones de la Ley de Benford esencialmente a la perfección, así como la media de los dígitos terminales se ajusta a un modelo uniforme con altísima correlación. Para investigadores como Kalinin y Mebane tal fenómeno es demasiado bueno para haber ocurrido naturalmente. Ellos argumentan que es la firma deliberada y descarada de los estafadores. En términos más generales, no existe una única prueba estadística que soluciones la controversias frente a todas las alternativas posibles.

### 3.4. Problemas de multiplicidad estadística

Cada vez que se aplica un conjunto de pruebas estadísticas a los mismos datos, la probabilidad de que al menos un resultado sea un «falso positivo» es generalmente mayor que la probabilidad de un falso positivo para cada prueba individual considerada individualmente. Este fenómeno se llama «multiplicidad estadística». Posiblemente, por causa del «sesgo de confirmación», la tendencia de la multiplicidad al aumentar la tasa de falsos positivos se exagera cuando se seleccionan las pruebas después de examinar los datos. Por ejemplo, si un investigador forense observa que los resultados con mayor frecuencia terminan en 0 o 5, y luego decide por esa razón probar si la aparición de 0 y 5 es anómala; entonces, la posibilidad de un falso positivo aumenta sustancialmente. Dadas las suficientes pruebas diferentes, es prácticamente

---

seguro que los datos fallarán en al menos una de ellas.

Para el análisis forense electoral usamos un conjunto de muchas pruebas para examinar los resultados electorales, la posibilidad de que al menos una prueba marque los resultados como sospechosos puede ser grande, incluso si la elección es «limpia». Como ilustración, supongamos que deseamos probar si el dígito terminal se comporta como si fuera aleatorio, en particular, si todos los dígitos tienen la misma probabilidad de ocurrir y si existe alguna conexión (o dependencia) entre los dígitos finales en diferentes grupos de informes. La pregunta estadística es si los dígitos finales están «distribuidos de manera independiente y uniforme en los dígitos  $\{0, 1, \dots, 9\}$ ». Hay una variedad de pruebas estadísticas estándar, pero son sensibles a diferentes tipos de anomalías. Por ejemplo, **1)** podríamos probar usando la media de los dígitos terminales (que se espera que sea 4.5 si los dígitos están efectivamente distribuidos uniformemente),<sup>[34]</sup> **2)** o usando la prueba de Kolmogorov de la distribución empírica contra la función de masa de probabilidad teórica que asigna la probabilidad 0.1 a cada uno posibilidad; **3)** <sup>[34]</sup> o usando una prueba de  $\chi^2$  para frecuencias iguales de todos los dígitos; **4)** o usando el prueba de rango multinomial para comparar los dígitos más frecuentes y menos frecuentes;<sup>[34]</sup> **5)** o una prueba de multimodalidad; <sup>[34]</sup>**6)** o una prueba basada en la frecuencia con la que ocurren los 0 y los 5; o cualquier número de otras pruebas. Si usamos suficientes pruebas de este tipo, y especialmente si examinamos los datos antes de elegir qué pruebas aplicar, puede ser bastante probable que al menos una clasifique las elecciones como «mancilladas».<sup>[32]</sup>

La desigualdad de Bonferroni es una forma conservadora de dar cuenta de la multiplicidad en las pruebas: el nivel de significación combinado (posibilidad de un falso positivo) de una colección de pruebas es como máximo la suma de sus niveles de significación, incluso cuando las pruebas tienen una dependencia estadística arbitraria (siempre que las pruebas consideradas se incluyen en la suma: la desigualdad de Bonferroni no protege contra falsos positivos si las pruebas se eligen después de analizar los datos). La posibilidad de una unión arbitraria de eventos es, como máximo, la suma de sus posibilidades separadas. Sin embargo, debido a que brinda protección universal, la desigualdad de Bonferroni puede ser bastante conservadora en algunas circunstancias.

---

### 3.5. Análisis de distribuciones numéricas de dígitos

El análisis forense electoral basado en dígitos (por sus siglas: AFEBD) intenta determinar si una elección fue fraudulenta observando los resultados numéricos en el nivel más fino en el que se informaron los recuentos de votos. Esto es particularmente atractivo debido a su bajo costo, pues, hoy en día, los recuentos de los votos locales suelen encontrarse disponibles en Internet. Los métodos están diseñados para clasificar un conjunto de declaraciones electorales como fraudulentas o libres de fraude al inspeccionar sus distribuciones numéricas y basarse en: **1)** suposiciones sobre la distribución de probabilidad de resultados en ausencia de fraude o malversación, y **2)** significado de pruebas de hipótesis nula.[34]

Los métodos AFEBD convencionales se basan en dos observaciones generales. Primero, muchos procesos naturales producen cantidades en las que los números se distribuyen de acuerdo con la ley de Benford. En 1995, Hill demostró que en los números extraídos de una mezcla aleatoria de distribuciones, sus números convergen en una distribución logarítmica. Según la Ley de Benford, en el sistema de números decimales, la distribución de números se aproxima a la uniformidad con el orden de los dígitos; es decir, para el cuarto dígito es prácticamente uniforme la distribución. La segunda observación es que los humanos (tal vez por cuestiones culturales) son pésimos generando números que sigan la Ley de Benford cuando se les pide que generen números aleatorios mentalmente. Combinando estas dos observaciones, esperamos que los números generados por elecciones precisas se ajusten a la Ley de Benford, mientras que los números inventados por estafadores distan de tal ley. Esta es la conjetura audaz que siguen los forenses electorales que estudian frecuencias de dígitos.[34]

De tal planteamiento, surgen dos líneas de crítica. La primera cuestiona la suposición de que los números en elecciones precisas deben seguir la Ley de Benford. Por ejemplo, las leyes y reglamentos pueden establecer límites en el número de votantes por mesa electoral (o barrio, recinto, etc.). Esto afecta las frecuencias de los números en el primer dígito de conteo de votos. Por esta razón, los métodos AFEBD convencionales generalmente se centran en el segundo dígito o en los dígitos de orden superior.

Eso sólo resuelve parte del problema. En el 2011, Shikano y Mack usaron simulaciones y datos de las elecciones federales alemanas, ellos afirman que incluso los resultados precisos pueden distar de la Ley de Benford, por razones distintas al fraude. Además, las irregularidades que no son causadas por fraude pueden producir números que no cumplen con la Ley de Benford.

---

Por ejemplo, los contadores de votos humanos cuyas habilidades matemáticas son débiles pueden favorecer ciertos dígitos (es decir: «redondear») para simplificar los cálculos, sin favorecer intencionalmente a ninguno de los candidatos. La segunda línea de crítica apunta a la suposición de que todas las irregularidades, o incluso todo fraude deliberado, resultarán en violaciones de la ley de Benford. A diferencia de los sujetos humanos en experimentos de laboratorio, algunos estafadores electorales saben qué características de los números serán inspeccionados, además de que cuentan con acceso a herramientas para generar números que pasarán prueba estadísticas basadas en la ley de Benford.

Una alternativa al AFEBD convencional implica «aprender» la distribución de resultados libres de irregularidades de los datos. Esto es especialmente atractivo si hay un conjunto de elecciones disponibles que se cree que son precisas, para elecciones que tuvieron lugar en condiciones similares a las de la elección bajo escrutinio. Esto permite relajar la fuerte suposición distributiva del AFEBD convencional: la distribución numérica libre de irregularidades no necesita ser conocida exactamente y no requiere ser la Ley de Benford. En cambio, la prueba se basa en una suposición más débil de que las distribuciones numéricas para elecciones precisas difieren de las de elecciones con irregularidades.

Hay dos estrategias para implementar tales métodos. Primero, uno puede tratar las elecciones que se consideran precisas como un conjunto de entrenamiento, estimar las características de la distribución de referencia a partir de ellas y comparar las elecciones evaluadas con esas estimaciones. En ese sentido, en el 2011, Cantu y Saiegh utilizaron datos sintéticos para obtener las características distributivas de (lo que ellos consideran) elecciones precisas. Alternativamente, uno puede evaluar si las elecciones precisas y las elecciones bajo escrutinio pueden ser adecuadas utilizando la misma distribución numérica. La segunda opción en efecto significa que la tarea ya no es clasificar si un conjunto de datos con irregularidades o no, sino más bien, si la distribución de los datos inspeccionados es «lo suficientemente cercana» a las que se cree que son precisas. Lo cual es la actitud que se espera de un perito en la corte.

### **3.6. Prueba de hipótesis nula basada en dígitos**

En AFEBD convencional, la distribución numérica observada se compara con otra distribución suponiendo que generó los resultados si no hubo irregularidades. La comparación conforma una prueba estadística de la hipótesis nula de que los datos extraídos de esa distribución hipotética. La prueba

---

de significación de hipótesis nulas (PSHN) se ha utilizado fructíferamente a problemas en una muy amplia variedad de dominios. Aplicar PSHN generalmente implica calcular un «índice estadístico de prueba» a partir de los datos, y luego, comparar su valor observado con la distribución de valores que tendría el índice estadístico de prueba si se aplicara a muchas muestras del mismo tamaño, suponiendo que la hipótesis nula es verdadera. Esta comparación se resume como un valor  $p$ , una medida de cuán inusuales serían los datos observados si la hipótesis nula fuera cierta. Los valores  $p$  más pequeños son evidencia más convincente de que la hipótesis nula es falsa.

El valor  $p$  no es la probabilidad de que la hipótesis nula sea verdadera, un error que se comete a menudo. Más bien, es la probabilidad de observar datos tan inusuales como los datos reales, calculados suponiendo que la hipótesis nula es verdadera. El último paso de la prueba de hipótesis es comparar el valor  $p$  con un umbral (predeterminado) de significancia estadística, como 5 %, 1 % o 0.1 %, para decidir si los resultados electorales reflejan irregularidades (que ocurre si el índice- $p$  es suficientemente pequeño).[34] Si la hipótesis nula es verdadera (es decir, si los resultados de la elección son precisos) y se consideran e inspeccionan muchas muestras repetidas, la hipótesis nula será rechazada falsamente por una fracción esperada conocida de muestras.[34] Para que el PSHN sea útil, entre otros requisitos, el investigador debe seleccionar el estadístico de la prueba y el umbral para el valor  $p$  para que se ajusten a la pregunta formulada de la prueba y la decisión que se alcanzará en base a ella. Esto es especialmente importante porque el «poder» de las pruebas<sup>3</sup> está limitada por el tamaño de la muestra.

En particular, las pruebas AFEBD pueden arrojar conclusiones diferentes para dos muestras en las que los números aparecen con las mismas frecuencias relativas, pero los tamaños de muestra difieren. Es más probable que la hipótesis nula sea rechazada para muestras de mayor tamaño si los datos tienen mismas densidades de probabilidad sobre dígitos y umbral de significancia. En principio, el nivel de significancia debe reflejar los costos y beneficios de las decisiones correctas e incorrectas (verdaderos y falsos positivos y verdaderos y falsos negativos). En la práctica, tales opciones de umbral informadas son raras.[34] En cambio, los profesionales generalmente eligen un nivel de significancia convencional, como 0.05, 0.01, 0.001 o 0.0001. Pre-especificar el nivel puede ayudar a garantizar que la elección no esté diseñada para ofrecer una decisión preferida. En el contexto del análisis forense electoral, tal vez sea más difícil de lo normal elegir un buen nivel, porque las elecciones generalmente se llevan a cabo en intervalos de años como eventos únicos,

---

<sup>3</sup>La capacidad de una prueba para concluir correctamente que la hipótesis nula es falsa cuando una hipótesis alternativa particular es verdadera.

---

dependiendo mucho del resultado y su legitimidad percibida. Esto es especialmente cierto para las democracias nuevas o problemáticas, que son más propensas a ser inspeccionadas para detectar signos de fraude electoral. Es imposible saber de antemano qué costos y beneficios se derivan de encender alarmas de fraude para elecciones precisas o de mantenerlas apagadas para elecciones con irregularidades.

### 3.7. Entre la precisión y la irregularidad electoral

En términos generales, PSHN busca clasificar los resultados electorales como precisos o irregulares, aunque la hipótesis nula requiere un modelo específico de cómo se generan resultados precisos. En la práctica, esta pregunta cualitativa es menos interesante que las cuantitativas: ¿cuánta irregularidad afectó los resultados?, ¿las irregularidades alteraron el resultado electoral? En el 2015, Medzihorsky propuso una clase latente para análisis forense basado en dígitos.[34] En este enfoque, los números observados se clasifican en dos clases: precisos e irregulares. Reciben el nombre de latentes en el sentido de ser inobservables, por lo que requieren ser deducidas por medio de los datos.

El AFEED convencional se fundamenta sobre el supuesto de que la forma exacta de la distribución numérica de resultados precisos se presume conocida *exprofeso*. Sin embargo, desconocemos la forma de la distribución numérica para resultados irregulares, más allá del hecho de que es diferente al estándar asumido.

El AFEED de clase latente descompone los resultados observados en el conjunto más grande posible (u otra distribución específica, como la uniforme) y el conjunto más pequeño posible. Proporciona una estimación de la fracción posible de resultados imprecisos.

En tal contexto, son posibles dos interpretaciones diferentes de la clase residual. La primera clase se compone de números que deben eliminarse para obtener resultados consistentes con la distribución de resultados. El segundo es la clase residual está compuesta de números que cambiarse para obtener una distribución de resultados precisos. La primera interpretación conduce a un caso especial de «índice de ajuste de la mezcla» [Rudas y colaboradores (1994)]. La segunda interpretación conduce a una medida relacionada con el índice de disimilitud [Gini (1914)].

El enfoque de clase latente no requiere la suposición convencional de que la distribución numérica de resultados precisos se conoce antes del suceso,

---

siempre que haya un conjunto de rendimientos que se cree que son precisos, además de los resultados examinados. En ese caso, el enfoque de clase latente clasifica los dígitos en un grupo que describe la mayor cantidad posible de todos los números observados y agrupa el resto en un grupo residual.

### 3.8. Conclusiones del capítulo

Existen dos reservas principales sobre los procedimientos estadísticos forenses al analizar los datos electorales para detectar irregularidades. Sin embargo, si los resultados electorales están disponibles, los análisis estadísticos pueden identificar patrones sospechosos y permitir el análisis de casos potenciales de fraude de una manera más integral que lo que logran las misiones de observación electoral. Identificando los datos adecuados, los métodos estadísticos pueden abarcar todo el territorio y llegar a cada centro de votación. Sin embargo, las anomalías estadísticas pueden ocurrir por muchas razones, no sólo por fraude o error, por lo que los resultados de pruebas estadísticas positivas nunca pueden considerarse una prueba directa de irregularidades, sino sólo evidencia que complementa otras evaluaciones cualitativas. Por ejemplo, a través de los observadores electorales. La manipulación de las elecciones o de sus resultados o la aparición de errores que alteran los resultados sólo se pueden probar a través de evidencia directa, como una auditoría rigurosa contra registros en papel confiables y verificados por los votantes. El uso masivo de tecnologías digitales para emitir el voto limitaría este tipo de estudios. Por el contrario, los procedimientos estadísticos pueden ignorar la existencia de algunos tipos errores graves y fraude, pues sólo son sensibles a una gama limitada de problemas dentro de los procesos electorales. Por lo tanto, no podemos concluir, a partir de la ausencia de anomalías estadísticamente notables, que las elecciones fueron bien administradas y libres de interferencia, manipulación o errores que alteren los resultados.

Así, identificamos cuatro limitaciones críticas para los procedimientos estadísticos como herramientas para detectar manipulaciones en las elecciones:

1. Muchos factores afectan los resultados electorales, incluso varían geográficamente, lo que puede generar patrones similares a los relacionados con el fraude. Las concentraciones de votos en regiones pequeñas, junto con una alta participación, pueden ser el resultado de la concentración territorial de minorías socioeconómicas o culturales, la prominencia y el poder de movilización de los candidatos locales o las comunidades pequeñas. Incluso en democracias consideradas como estables, la participación tiende a correlacionarse con el voto partidista (Grofman et al., 1999). Las distribuciones desiguales de dígitos pueden resultar del

---

diseño de distritos electorales, estrategias de nominación de candidatos o alterar las estrategias de votación entre los distritos electorales (Mebane, 2013). Además, los «errores inocentes» de los administradores electorales o los trabajadores electorales pueden ser indistinguibles del fraude por métodos estadísticos.

2. Las herramientas estadísticas para el análisis forense electoral siempre se basan en pruebas probabilísticas. Es improbable que los patrones que señalan como anómalos ocurran al azar de acuerdo con su supuesto modelo de elecciones precisas, pero pueden ser posibles. Si se aplica la misma prueba estadística a un gran conjunto de elecciones que fueron todas precisas, deberíamos esperar que algunas de las elecciones se marquen como sospechosas, porque los patrones sospechosos surgen por casualidad. Del mismo modo, si se realizan muchas pruebas diferentes para diferentes tipos de irregularidades para las mismas elecciones, es probable que una de estas pruebas muestre un «falso positivo».
3. El proceso electoral tiene muchos pasos, por ejemplo, registro de candidatos, compilación del registro de votantes y conteo de votos. Cada paso puede sufrir irregularidades (en diferentes formas). Por lo tanto, las herramientas estadísticas son incapaces de probar sólo una forma o fuente de irregularidad. Si bien existen pruebas sofisticadas para algunas instancias y formularios, las pruebas para otras son mucho más rudimentarias, y para muchos pasos en el proceso electoral existe carencia de datos sistemáticos o pruebas disponibles.

Una vez más, las pruebas estadísticas siempre se basan en suposiciones probabilísticas sobre cómo se generan resultados precisos y, por lo tanto, las irregularidades pueden pasar desapercibidas por casualidad o porque no son lo suficientemente penetrantes o lo suficientemente grandes como para ser detectadas. Por ambas razones, las pruebas estadísticas múltiples que no marcan ningún patrón sorprendente nunca deben interpretarse como evidencia afirmativa de que las elecciones son precisas.

4. Un estafador sofisticado puede adaptarse a los procedimientos estadísticos estándar, adoptando contramedidas para asegurarse de que su trabajo pasará las pruebas. Por ejemplo, al aleatorizar algunos de los números manipulados, las pruebas estadísticas basadas en distribuciones numéricas ya no identificarán estos resultados como irregulares. En todo caso se espera que progrese el estudio de auditoría forense para combatir tal práctica.

---

Es posible detectar irregularidades electorales por medio del análisis estadístico de la base de datos de resultado, pero se debe ser cauto en la interpretación de los resultados.

## Capítulo 4

# Propuesta de análisis y auditoría

Si se decide utilizar la ley de Benford para analizar una base de datos. ¿Cómo se debe calibrar?

Creemos que desde el punto de vista metodológico es inadecuado utilizar una base de datos de un proceso electoral pasado (aunque cuente con alta reputación pese que goce de prestigio de bien realizada y limpia). Debido a que la mayoría de los analistas de bases de datos electorales –entre ellos el Dr. Mansilla– usan una base de datos de referencia (*e.g* elecciones del año 2006) que en todos los casos fue construido por personas con intereses diversos y poco objetivos ¿Cómo sabemos que está base está bien construida?. Una afirmación como: «Si no hubo errores en el conteo electoral debería de seguir perfectamente a la Ley de Benford» carece de sentido pues desconocemos el mecanismo intrínseco que lleva a la gente a votar y que luego se refleja en los resultados en un análisis estadístico de dígitos. Lo que tiene más sentido es afirmar que: «Si esta elección esta limpia de error debería de tender a la ley de Benford»; pero, cual es el límite numérico, en que momento se debe encender una alarma de sospecha de fraude. La mayoría de los autores establecen su propio umbral numérico. Primero seleccionan una base de datos como limpia, después otra como mancillada, de donde presentan su propio umbral. El principio de autoridad de tales fuentes de información debe ser invalido cuando se trata de construir un instrumento de análisis objetivo.

Por ello, es mucho mejor que la base de datos de calibración provenga de una fuente que claramente no fuera manipulada subjetivamente por un humano. Si bien los datos que provienen de mediciones pueden ser una opción, pues siguen la ley de Benford (como se expuso en la sección 2.1), es mucho mejor utilizar un sucesión matemática para calibrar la base de datos. Pues en ella son más claros las construcciones (de tamaño y alteración controlada) impuestas al manipular la base de datos. En los siguientes párrafos expone-mos como es que la secuencia de Fibonacci puede ser útil para comprender

---

mejor el significado operativo de la ley Benford para analizar una base de datos.

## 4.1. Criterios de ajuste entre modelos teóricos y datos empíricos

A continuación expondremos –de modo breve– los fundamentos y las características principales de cuatro índices estadísticos que utilizamos en esta tesis.

### 4.1.1. Desviación media absoluta MAD

La desviación absoluta media (o por sus sigla en inglés: MAD), también conocida como «desviación media» es la media de las desviaciones absolutas de los datos en torno a la media de los datos. En otras palabras la distancia promedio (absoluta) respecto de la media. La «desviación absoluta promedio» puede referirse a este uso, o a la forma general con respecto a un punto central especificado. En términos matemáticos, para el análisis MAD del primer dígito lo podemos escribir como:

$$\text{MAD} = \sum_{i=1}^9 \frac{|\text{Observado}(i) - \text{Teoría}(i)|}{n}. \quad (4.1)$$

Algunos investigadores han propuesto que MAD se utilice en lugar de la desviación estándar, ya que corresponde mejor a la vida real [3]. Debido a que MAD es una medida de variabilidad más simple que la desviación estándar, puede ser útil en la enseñanza escolar básica y en particular puede ser útil para los profesionales forenses y del Derecho que cuentan con escasos conocimientos estadísticos [4] [5].

La precisión del pronóstico de este método está muy relacionada con el método del error cuadrado medio (en inglés, *mean squared error*: MSE), que es sólo el error cuadrático medio de los pronósticos. Aunque estos métodos están estrechamente relacionados, MAD es más comúnmente utilizado porque es más fácil de calcular (evitando la necesidad de elevar al cuadrado) [6] y es más fácil de entender por ser una operación más sencilla [7].

---

#### 4.1.1.1. Interpretación de MAD

Por ser una diferencia entre lo observado y la teoría (en nuestro caso la ley de Benford) mientras más pequeño sea el índice MAD, mejor. Pero, ¿cuándo el MAD es lo suficientemente grande como para señalar una sospecha de fraude en un conjunto de datos?

Recordemos, en ciencia forense, los resultados de una pericial deben ser presentados a un abogado, juez o tribunal constituido por personas; de modo que se requiere una traducción del significado de los intervalos numéricos. Al ser MAD una operación relativamente simple, se propone emplearla para detectar fraude en bases de datos contables.

En lo que nos atañe, de acuerdo con el libro «*Digital Analysis Using Benford's Law*», M. Nigrini sugiere el criterio mostrado en la Tabla 4.1. Es importante resaltar que los umbrales, como en la transición de «aceptable» a «marginal», pueden variar de acuerdo con la naturaleza y tamaño de la muestra de la población. Por ejemplo, si se considera una base de datos de una población, en nuestro caso de resultados electorales. Y usamos un software para extraer el primer dígito de dicha base de datos entonces procedemos a realizar un contraste con el modelo teórico (es decir, la Ley de Benford) con respecto a dicha base. Usamos el MAD como criterio para observar si se ajusta a dicha ley. Suponemos que nos arroja un valor de 0.001. En éste caso, el procedimiento a seguir es el siguiente. Se revisa la Tabla 4.1 Se verifica a que valor «categórico» corresponde a dicho valor numérico. En nuestro ejemplo, se puede observar que el ajuste a la Ley de Benford es «Cercano». es decir, la base de datos estudiada es una muy buena aproximación a la ley de Benford. Nuestra alarma se «encendería» cuando el índice MAD supere 0.012 sin importar la naturaleza de la base.

En la literatura, la mayoría de los ejemplos o propuestas de estandarización consisten en información obtenida de actividades o constructos humanos. Por ejemplo, Mark Nigrini en su libro «*Digital Analysis Using Benford's Law*» propone usar la Ley de Benford para varias bases de datos construidas por personas. Entonces ¿como sabemos si está bien calibrado ese criterio? ¿Cómo

Tabla 4.1: Criterios de interpretación de MAD, de acuerdo con M. Nigrini.

| Ajuste     | Intervalo     |
|------------|---------------|
| Cercano    | 0.000 - 0.004 |
| Aceptable  | 0.004 - 0.008 |
| Marginal   | 0.008 - 0.012 |
| Inadecuado | >0.012        |

---

sabemos si los umbrales propuesto están bien contruidos? Sin embargo, si buscamos formalidad y rigor en el tema; entonces, una mejor propuesta es que el estándar sea una serie matemática, pues contará con alta correlación con la ley Benford <sup>1</sup>. Pero para que también cuente un contexto en las experiencias fácticas (actividades humanas, como la electoral) tal serie debe ser ampliamente utilizada para describir fenómenos naturales, como puede ser la sucesión de Fibonacci.

En consecuencia, pretendemos obtener un estándar que permita unificar criterios en el uso de índices estadísticos y en su presentación ante tribunales.

En otras palabras, lo que proponemos es usar un método que consiste en proponer la sucesión de Fibonacci como ente matemático y calcular su MAD con respecto a la distribución teórica, en éste caso la ley de Benford. Después contrastaremos los resultados con la prueba  $\chi^2$  y la prueba Z. Esto con la finalidad de encontrar una equivalencia entre pruebas al comparar una base de datos y la teorización de la ley Benford.

#### 4.1.2. Prueba ji-cuadrada

Tradicionalmente se suele clasificar la estadística en descriptiva e inferencial. La primera se asocia a la presentación de tablas, gráficas y mediciones que describen en forma general el comportamiento de una muestra o una población. Por otra parte, la inferencia estadística es asociada con la labor de investigación científica en la búsqueda de relaciones causales u asociaciones entre fenómenos diversos. En sus inicios, la investigación y desarrollo en torno a la estadística inferencial se orientaron en los métodos llamados «paramétricos», tales métodos de estimación y procedimientos de pruebas de hipótesis se apoyan a una serie de supuestos para su correcta aplicación y obtención de conclusiones certeras. Los científicos especialistas en ciencias que requieren el uso de inferencia estadística, pero cuyas observaciones no siempre cumplen con los supuestos necesarios para la aplicación de las pruebas clásicas, tales como la normalidad, la homogeneidad de varianzas y la exactitud de las mediciones, motivaron a la investigación de estadísticos y matemáticos en el terreno de procedimientos inferenciales que no plantean tantos requisitos y que sin embargo, coadyuvaran a conclusiones científicamente válidas.

La estadística no-paramétrica se desarrolló rápidamente a que sus procedimientos cumplen en general con las siguientes ventajas:

1. Los métodos no-paramétricos hacen pocos supuestos acerca de las poblaciones a partir de las cuales se obtienen las observaciones.

---

<sup>1</sup>Debido a que un buen indicador de aproximación a la ley de Benford sería ver su correlación cercana a 1, que es lo ideal y valor máximo.

- 
2. En general son fáciles de aplicar y los paquetes de análisis estadístico más populares incluyen opciones no-paramétricas en sus menús.
  3. Se pueden aplicar en situaciones donde los métodos tradicionales no son factibles de aplicación, tales como aquellas en las que las observaciones no son mediciones en el sentido preciso de la palabra, sino solamente intervalos de decisión. Como puede ser el caso de declarar limpia o mancillada una elección presidencial.

El desarrollo de la inferencia estadística no-paramétrica (cuyo nombre más adecuado sería inferencia estadística de distribución libre) se inició vigorosamente por la tercera década del siglo pasado. R. Savage señala el año de 1936 como el punto de arranque debido a la publicación de un artículo sobre correlación por rangos debidos a Hotelling y Pabst. Otros autores como Scheffé identifican sus antecedentes en fechas anteriores. En el periodo de 1900 a 1911 se publican algunos artículos debidos a Karl Pearson y años después se presenta la prueba de los signos en la primera edición de *Statistical Methods for Research Workers* (1925) escrita por Ronald A. Fisher. Por el año de 1943 Scheffé escribió: «Solamente una pequeña fracción de la extensa literatura dedicada a la estadística está orientada a caso no-paramétrico y la mayor parte ha sido escrita en ésta década. Esperemos, sin embargo, un rápido crecimiento para esta rama. La perspectiva de una teoría libre de supuestos específicos acerca de la forma de las distribuciones de las poblaciones deber excitar la mente tanto de teóricos como prácticos y que la teoría combina la elegancia estructural con posibilidades de aplicación».

En 1947, F. Wilcoxon promueve los métodos estadísticos no-paramétricos con la publicación de un libro titulado *Some Rapid Approximate Statistical Procedures*, Para el año siguiente M.G.Kendall publica el primer libro orientado a métodos apoyado en intervalos. *Rank Correlation Methods* Kendall cita las posibilidades de aplicación de tales métodos en campo tan diversos como la psicología, educación, economía y experimentación industrial. Alrededor de 1956, Sidney Siegel publica el que posiblemente sea el texto más popular de métodos estadísticos no-paramétricos de la actualidad: *Non parametric Statistical for the Behavioral Sciences*.

Para 1962 el impulso de la estadística no-paramétrica es tal que Savage publica sus *Bibliography of Nonparametric Statistical* con referencia a más de 3000 artículos relacionados con el tema. J. Háyek uno de los más notables investigadores de ésta rama de la estadística escribe en 1969: «Los métodos no-paramétricos constituyen una de las ramas más exitosas de la estadística moderna: Ellos son ampliamente aplicables, de rápida ejecución y fáciles de entender: La teoría no-paramétrica es elegante y dos décadas de vigoroso

---

desarrollo no han agotados sus posibilidades de estimular a los investigadores».

En concreto, sobre la prueba ji-cuadrada (o  $\chi^2$ ), esta es una prueba de bondad de ajuste (descubierta por Karl Pearson) que sirve para contrastar una prueba de hipótesis referida con respecto a su distribución de frecuencias. En general, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula. Como representación matemática podemos escribir:

$$\chi^2 = \sum \frac{(\text{Observado} - \text{Teoría})^2}{\text{Teoría}} \quad (4.2)$$

El uso de  $\chi^2$  sirve para probar la asociación entre dos variables utilizando una situación hipotética y datos de una fuente empírica o teórica o combinaciones de estas. Por ejemplo, podemos describir como una variable el bienestar personal y como segunda variable la realización de un deporte. Por tanto, al asociar éstas dos variables con la prueba  $\chi^2$  y una muestra podemos conjeturar<sup>2</sup> que pueden ser variables dependientes una de otra o simplemente ser independientes.

Otro ejemplo de aplicación sería el siguiente. Supongamos que tenemos 57,000 transacciones financieras de ingresos de IBM relacionadas con el primer trimestre de 2008. Entonces, estamos tratando de investigar si IBM dio o no información honesta; por tanto, se estudia la distribución de primer dígito en comparación con la distribución logarítmica. En consecuencia, concluimos con respecto a los informes si son veraces o fraudulentos. Entonces, éste estudio se llega a través de una prueba  $\chi^2$ . Para aplicar éste método se realiza el siguiente procedimiento.

1. Se escoje una muestra aleatoria de tamaño  $n$  de la población cuya distribución de probabilidad es desconocida.
2. Ordenamos las  $n$  observaciones en  $k$  intervalos de clase, denotamos por  $O_i$  la frecuencia observada en el intervalo de clase  $i$ .
3. Calculamos la frecuencia esperada a partir de la distribución hipotética.
4. Para el intervalo de clase  $i$ -ésimo, denotado por  $E_i$ , el índice estadístico de prueba es calculado mediante la Ec. 4.2.

---

<sup>2</sup>Correlación no implica causalidad.

---

### 4.1.3. Prueba-Z

La prueba Z es una prueba de hipótesis basada en el estadístico Z, que sigue la distribución normal estándar sujeto la hipótesis nula.

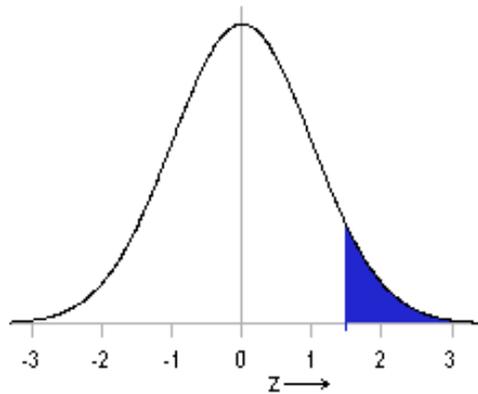


Figura 4.1: Esquema de los elementos de la prueba Z, la curva (línea negra) representa la función de densidad y el área bajo la curva (color azul) representa la probabilidad acumulada.

La prueba Z más simple es la prueba Z de una muestra, la cual evalúa la media de una población normalmente distribuida con varianza conocida.

Para aplicar ésta prueba se plantea el siguiente procedimiento:

1. Se plantea la hipótesis nula  $H_0$  y la hipótesis alternativa  $H_1$
2. Seleccionamos el nivel de significancia  $\alpha$  y el tamaño de la muestra  $n^3$ .
3. Determinamos el índice estadístico de prueba y la distribución muestral apropiados.
4. Determinamos los valores críticos de la zona de rechazo y no rechazo.
5. Recopilamos los datos y obtenemos el índice estadístico de prueba mediante la ecuación siguiente:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (4.3)$$

donde  $\bar{x}$  es el promedio,  $\mu$  es la media teórica y  $\sigma$  es la desviación estándar y  $n$  el tamaño de la muestra.

---

<sup>3</sup>Si el tamaño de la muestra es menor a 30 entonces se utiliza la prueba t de Student.

---

6. Finalmente, tomamos la decisión estadística y sacamos las conclusiones.

Nigrini en su obra menciona que la prueba-Z puede usarse para probar si un muestra grande y aleatoria sigue o no la distribución de Benford. El plantea como hipótesis nula como lo siguiente:

H<sub>0</sub>: el conjunto de datos sigue la ley de Benford. H<sub>1</sub>: El conjunto de datos no sigue la ley de Benford.

#### 4.1.4. Prueba-Z para diferencia de proporciones

Esta prueba se basa en la aproximación normal de la distribución binomial. Queremos comparar dos proporciones,  $p_1$  y  $p_2$ , observadas en dos grupos distintos de tamaños  $n_1$  y  $n_2$ , respectivamente. Esta prueba es utilizable cuando los tamaños muestrales  $n_1$  y  $n_2$  son grandes, para poder aplicar el Teorema Central del Límite. El estadístico de contraste se calcula como:

$$z = \frac{P_1 - P_2}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} \quad (4.4)$$

[35]

#### 4.1.5. Análisis por regresión lineal

En 1886, Francis Galton escribió que aunque los padres altos engendran generalmente hijos altos, la estatura de éstos tiende a ser menor; mientras que los padres bajos suelen tener hijos bajos, la estatura de éstos tiende a ser mayor que la de sus padres. Es decir, existe una tendencia a «regresar» a la estatura media poblacional. Documentalmente, fue la primera vez que el término «regresión» se utilizó con fines estadístico descriptivos. Tal enunciado conocido como la «Ley de Regresión Universal de Galton», fue confirmada por Karl Pearson al analizar más de 1000 casos [36].

Actualmente, el análisis de regresión es una técnica estadística que trata de la relación de una variable dependiente en función de una o más variables independientes. Pero también se puede utilizar para comparar la concordancia con una teoría y datos observados.

El proceso es sencillo. Consiste en emparejar unívocamente los valores predichos por una teoría (*e.g.* variable independiente en el eje abscisa) con los datos observados empíricamente (variable dependiente en el eje ordenada). Si existe una alta concordancia entre la teoría y la observación, entonces la mejor línea de ajuste los datos tenderá a ser la línea recta identidad:  $y = 1x + 0$ .

---

En efecto, si existe concordancia entre teoría y observación; entonces, dato a dato debe existir similitud. Por lo que la pendiente de la mejor línea de ajuste será cercana al valor 1. Así, el valor de la variable independiente debería ser cero. Por supuesto, es de esperar que un valor de correlación también sea cercano a 1. Cuando no se cumplan estas tres condiciones se debe revisar el contexto en que se realizaron las observaciones para explicar las desviaciones. Por ejemplo, si estas variaciones implican un tipo de error sistemático que desvía a la pendiente del valor unidad, de calibración que desplaza al valor de la variable independiente, o un excesivo ruido que altera la correlación. O bien, otro tipo de error que se puede presentar en la comparación entre un modelo teórico y una observación empírica. De este modo, este tipo ha sido aplicado en varios trabajos académicos.

Siendo la regresión lineal una herramienta estadística, por convención, esta empieza a ser confiable cuando se comparan diez parejas de datos teórico-empíricos. Así, los análisis con nueve datos (como los estudios de la aparición del primer dígito en bases de datos) son únicamente orientadores. Más aún, la obtención de valores ideales no implican relación causa-efecto. Pero el caso contrario, la falta de congruencia entre los datos, implica que se debe revisar las fuentes de información o el plantamiento del problema de investigación. En este trabajo, mostramos los alcances de esta perspectiva para analizar la concordancia entre la ley del primer dígito y bases de datos electorales.

A la correlación que nos referimos en este texto es:  $R^2$ , en la literatura especializada se le llama coeficiente de determinación, que para el ajuste lineal y solamente en este caso coincide con el cuadrado de la correlación de Pearson:  $r$ . Si representamos por  $\sigma^2$  a la varianza de la variable dependiente y la varianza residual por  $\sigma_r^2$ , el coeficiente de determinación se puede escribir como:

$$R^2 = 1 - \frac{\sigma_r^2}{\sigma^2} \quad (4.5)$$

$R^2$  es un índice normalizado, su peor y mínimo valor es cero, mientras que su valor mejor y máximo es 1. Dependiendo de la cultura científica, dentro de una especialidad, el umbral de una correlación aceptable cambia. Al parecer, en la mayoría de los trabajos en ciencias físicas que reportan  $R^2$ , deben superar 0.9. En contraste, trabajos en áreas biológicas y sociales pueden aceptar valores incluso inferiores a 0.5. Al parecer el criterio de umbral se debe a una interpretación sobre el alcance de  $R^2$ . Este índice lo podemos escribir como:

$$R^2 = \frac{\text{Variación(promedio)} - \text{Variación(recta)}}{\text{Variación(promedio)}}, \quad (4.6)$$

donde Variación(promedio) se refiere a la variación de los datos a un ajuste al valor promedio de los datos, que es una constante. Es decir el promedio de los datos formando una línea recta horizontal:  $y = \bar{x}$ . Por su parte, Variación(recta) es la variación de los datos alrededor de la línea ajustada; para nuestro caso ideal:  $y = 1x + 0$ . La Fig. 4.2 representa esta idea.

Cuando la variación alrededor de la línea es pequeña  $R^2$  tiende a 1, pero cuando las variaciones alrededor de la línea son grandes, estas tienden al valor promedio, de modo que  $R^2$  es cercano a cero. Por ejemplo si  $R^2 = 0.8$  implica que existe una variación del 80 % entre los datos y la recta ajustada. Además, se puede afirmar que la relación de la línea recta ajustada puede explicar el 80 % de la variación en los datos y que son otras las relaciones y variables que pueden explicar el 20 % de las variaciones entre datos. Si se trata de una sola relación que explicara por completo ese 20 %; entonces, deberíamos reconsiderar el modelo. Sin embargo, si son millones las relaciones que explicarían esa variación del 20 %; entonces, nuestra relación original de la recta es un excelente ajuste.

Es posible que en especialidades donde se intuye la existencia de muchas otras variables y relaciones los umbrales de correlación se acepten más bajos que en campos donde se conjetura que son pocas tales relaciones y variables. En definitiva,  $R^2$  no es la probabilidad de estar en lo correcto, es una medida de ajuste desde una perspectiva.

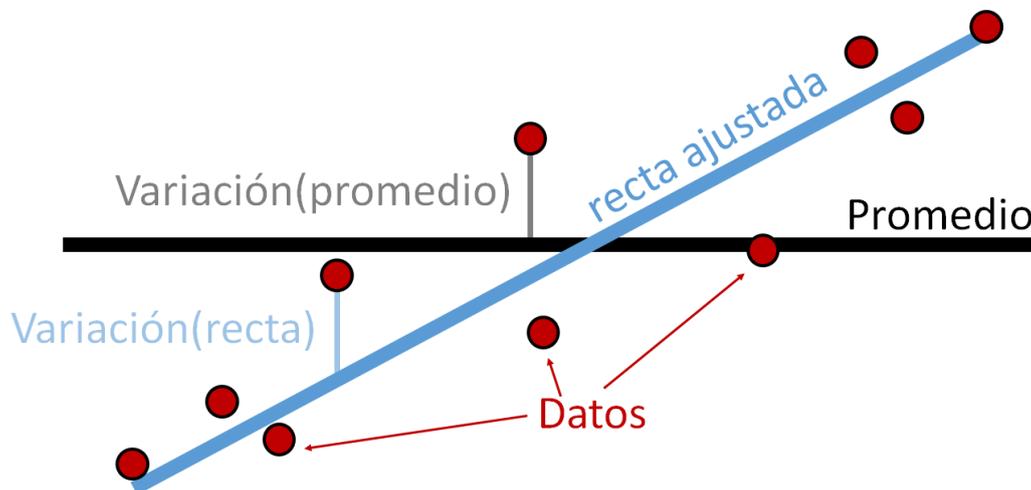


Figura 4.2: Interpretación para  $R^2$  en función de las variaciones de datos.

---

## 4.2. Cuarteto de Anscombe

Tanto para el análisis por regresión lineal, como otras propuestas, es importante primero inspeccionar la distribución espacial de los datos. Es decir, observar la gráfica. Porque diferentes conjuntos de datos pueden compartir las mismas propiedades estadísticas. Tal es el caso del llamado: «cuarteto de Anscombe».

Este cuarteto consiste de cuatro conjuntos de datos cuyos gráficos evidentemente son diferentes, pero que comparten la media las variables en  $x$  y también  $y$ , la varianza tanto en  $x$  y  $y$ . La correlación del par  $xy$  y la ecuación de la recta de regresión. Los datos son afectados por ruido, pero solo donde la línea recta es una buena descripción, en otro caso existe correlación pero no es lineal. En los otros dos casos, basta un sólo dato atípico para alterar la estadística para conjuntar el cuarteto.

En esta tesis, tratamos de evitar las meta-conclusiones a partir de solo observar índices estadísticos. En su lugar, creemos la necesidad de graficar y obtener índices de ajuste, pues implican una relación entre ellos. Por ejemplo, un buen ajuste a la línea recta debe implicar una alta  $R^2$  lo que abre el camino para analizar la recta ajustada: la pendiente de la recta  $m$  y la ordenada al origen  $b$ . Así, tres índices estadísticos se relacionan y se puede comparar mediante figuras. Tradicionalmente, relacionar otros índices como  $X^2$  y  $p$  es más laborioso.

Por todo lo mencionado, proponemos utilizar la sucesión de Fibonacci alterada como un punto común para relacionar los índices estadísticos y facilitar su interpretación en un dictamen o frente a una corte.

## 4.3. Sucesión de Fibonacci

En matemáticas, la sucesión o serie de Fibonacci, hace referencia a la secuencia ordenada de números descrita por Leonardo de Pisa, matemático italiano del siglo XIII. La secuencia con los siguientes números: 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144,...

La sucesión de Fibonacci comienza con los números 0 y a partir de estos, cada término es la suma de los dos anteriores. de forma general la regla de recurrencia se puede escribir como:

$$f_n = f_{n-1} + f_{n-2} \quad (4.7)$$

a partir de los valores predeterminados  $f_0 = 0$  y  $f_1 = 1$ . La Fig. 4.4 presenta la gráfica de la sucesión hasta  $f_{20}$  y el ajuste exponencial de esta serie de

potencias. Es claro que su comportamiento la hace un sucesión adecuada para estudiar la ley de Benford.

La sucesión de Fibonacci cuenta con aplicaciones en diferentes campos del conocimiento pues parece que modela de modo adecuado diferentes fenómenos relacionados con procesos biológicos y complejos. Ejemplos de su utilización los encontramos en ciencias de computación, matemática, y teoría de juegos. También encontramos usos del la sucesión de Fibonacci en área de finanzas al tratar de medir y pronosticar las alzas y bajas de los precios de las acciones. En concreto, se establecen ciclos, dónde el precio fluctúa dentro de patrones que se pueden, de cierta manera, predecir utilizando elementos que ofrece la sucesión se Fibonacci. Los economistas e inversores están utilizando software con medidas que se obtienen de Fibonacci para este efecto[37] .

Esta sucesión matemática, con los suficientes elementos, tiende a seguir a la ley de Benford. Aquí no proponemos utilizar esta sucesión como modelo de una elección limpia. Al contrario, nuestra sugerencia es: a partir de la sucesión de Fibonacci obtener un conjunto de números que tienda a la ley de Benford, después a ese conjunto alterarlo para que sea claro que ya no sigue a la ley del primer dígito, luego analizar estadísticamente esa sucesión descompuesta, de tal modo que sus índices estadísticos se conviertan en el

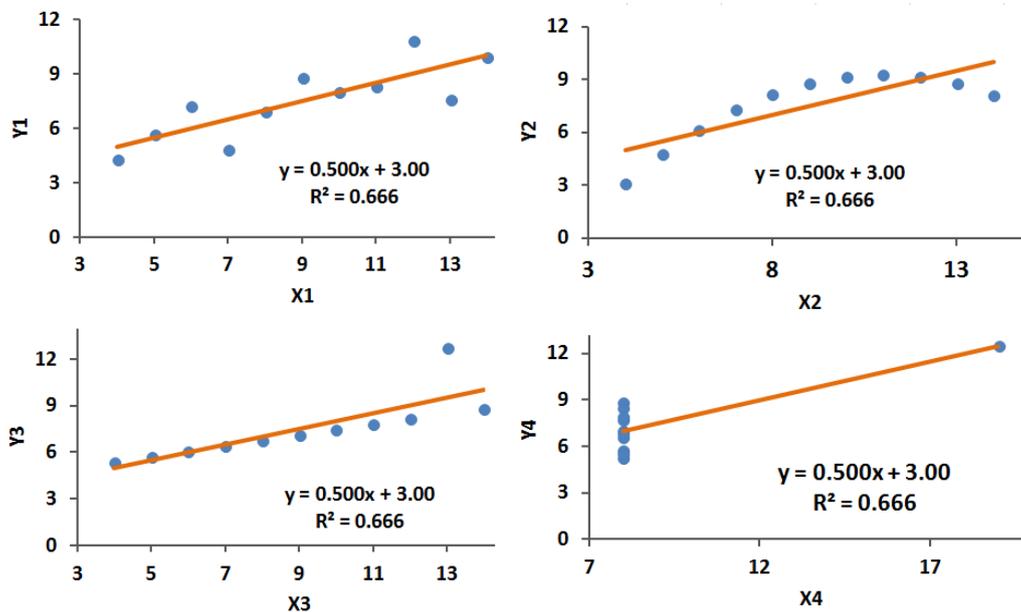


Figura 4.3: Gráficas del cuarteto de Anscombe. Estas cuatro distribuciones son muy diferentes entre sí, pero sus índices estadísticos son iguales hasta la cuarta cifra significativa. Imagen Vía: GPL: Wikimedia.

---

estándar de explicación y umbral numérico al estudiar otras bases de datos, en particular las electorales.

#### 4.4. Calibración de índices estadísticos a partir de la serie de Fibonacci

Proponemos utilizar la sucesión de Fibonacci como referencia de calibración debido a que es un objeto matemático carente de la subjetividad de bases de datos calificadas por limpias. Por ejemplo, Nigrini y otros autores para justificar un criterio umbral muestran bases de datos arquetípicas tanto limpias como mancilladas. ¿Cómo sabemos que tales datos son intrínsecamente mancillados cuando carecemos de una teoría sólida de como deben comportarse los datos carentes de manipulación? Tal vez, la respuesta es cambiar de perspectiva, en lugar de modelar la base de datos limpia, puede ser mejor idea establecer una base que ha dejado de seguir la ley de Benford debido a una manipulación objetiva, intencionada y clara.

Lo primero es establecer una punto base donde la sucesión de Fibonacci

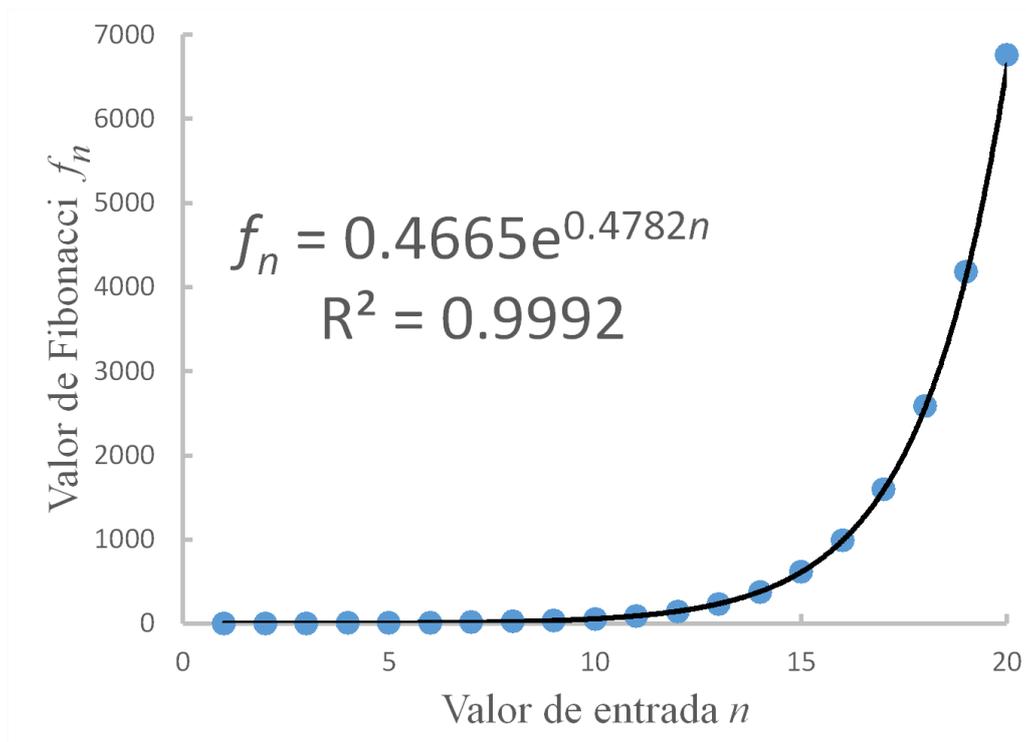


Figura 4.4: Gráfica y ajuste exponencial a la sucesión de Fibonacci hasta  $f_{20}$ .

tienda a ley de Benford. Es evidente que los dos primeros elementos de la serie no siguen la ley de primer dígito, por lo que se debe establecer un punto inicial. Escogemos como criterio de que la serie y ley de Benford siguen el mismo camino. en principio puede ser cualquiera, aunque lo preferible en esta etapa es escoger un criterio sencillo de explicar y popular. Así, escogemos el criterio MAD para establecer el umbral. Después revisamos desde cual valor  $n$  es posible que Fibonacci siga a ley de Benford La tabla 4.2 resumen los umbrales para la sucesión de Fibonacci. Encontramos que el valor  $n = 152$  es un umbral entre el criterio la «aceptable» a «cercano». Valores superiores a

Tabla 4.2: Mejora del criterio MAD aplicado a la sucesión de Fibonacci en función del número de elementos de la sucesión  $N$ . El umbral es la sucesión número 152; pues el criterio MAD sube de «aceptable» a «cercano».

| $N$        | MAD               | Ajuste           |
|------------|-------------------|------------------|
| $10^1$     | 0.05959           | No aceptable     |
| 43         | 0.0137508         | Marginal         |
| $10^2$     | 0.00636           | Aceptable        |
| <b>152</b> | <b>0.00654615</b> | <b>Aceptable</b> |
| $10^3$     | 0.000819877       | Cercano          |
| $10^4$     | 0.0003607         | Cercano          |

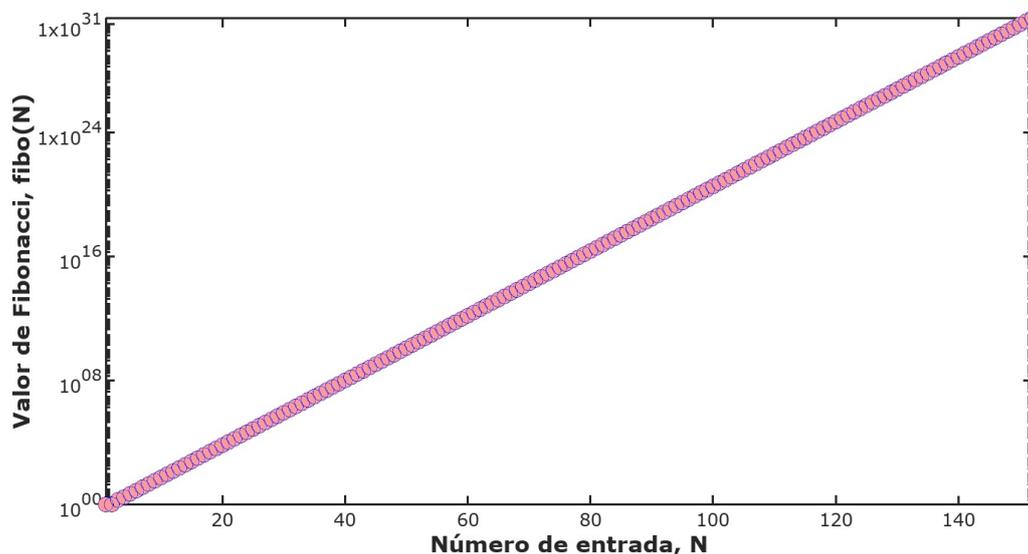


Figura 4.5: Los primeros 152 valores de la sucesión de Fibonacci, graficados en escala logarítmica.

---

152 son considerados cercanos a ley de Benford. Entonces valores mayores de 152 pueden servir para proponer la futura base datos mancillada. La Fig. 4.5 exhibe como estos 152 datos se comportan como lo hace una función de potencias; pues si bien el eje de las abscisas es lineal, el eje de las ordenadas es no es lineal, es logarítmico.

### Cómo debemos alterar la sucesión de Fibonacci

Siguiendo la idea expuesta en la sección 4.2 sobre el cuarteto de Anscombe, el cambio debe ser significativo, incluso en una gráfica. Los dígitos 1 y 9 son los poseen los valores máximo y mínimo en la frecuencia de la ley del primer dígito. Por tanto, intercambiar sus frecuencias es la alteración más notable. Así, todos los valores de 1 en el primer dígito se transforman en 9. La alteración entre los dígitos intermedios es innecesaria, pues son menos significativos y lo que deseamos es mancillar la base de datos de un modo claro. Este proceso se ilustra en la Fig. 4.6

### ¿Cuántos datos se deben alterar en esta sucesión?

Si bien  $f_{152}$  sigue la ley de Benford a un grado de etiquetarlo como «aceptable» en la perspectiva de MAD, esta colección de números carece de suficientes primeros dígitos 1 para cambiar a la etiqueta a «No aceptable». Una calificación más alta no sirve para nuestros propósitos de modelar una base de datos que siga a la ley de Benford y que luego de su alteración sea claro que es discordante con tal teoría. Entonces, debemos aumentar en la sucesión

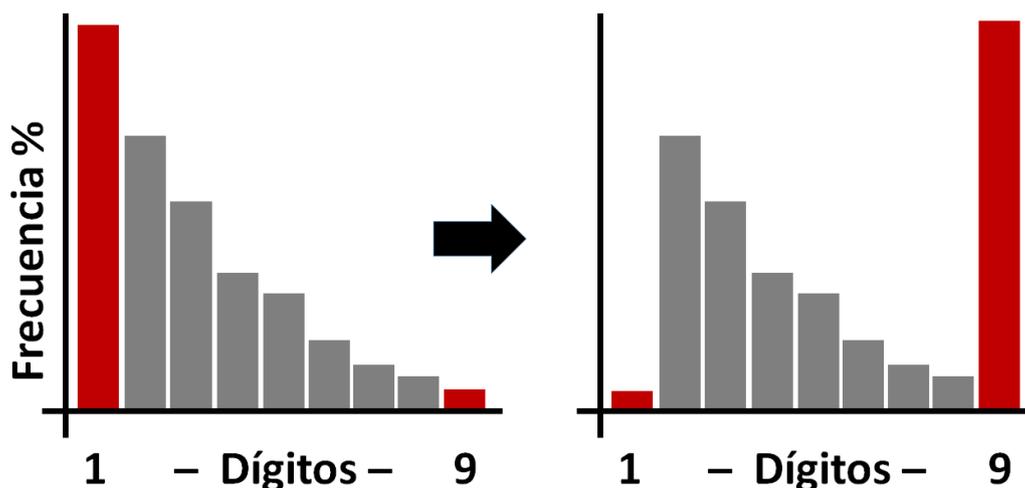


Figura 4.6: Ilustración del método de sustitución de 1 a 9 utilizado.

de Fibonacci e intercambiar 1 a 9 en la primera posición para alcanzar la meta. De este modo, proponemos que ese umbral sea  $f_{175}$ : con solo 13 cambios en la base de datos.

Según el criterio de Nigrini, al hacer 13 sustituciones de 1s por 9s de una muestra de 175 dígitos de la sucesión de Fibonacci nos degradará a la etiqueta «no aceptable»; es decir, que a partir de éste momento la muestra está manipulada; es decir, deja de seguir a la ley Benford. Qué en esta etapa, es lo buscado.

El modelo propuesto es más robusto, debido que se está usando una muestra objetiva y matemática, y nos permite considerarlo como punto de referencia para comparar con otras muestras o bases de datos. Ya que sabemos con certeza su origen y construcción puesto que es un objeto matemático y no está construida por una persona con intereses latentes, inobservables u ocultos. Por lo tanto, al aplicar el MAD a otra muestra y compararla con la de Fibonacci, contaremos con un criterio más firme para decidir profundizar en la investigación de un posible error o delito.

Éste criterio permite a la parte académica en la ciencia forense a tener una herramienta más poderosa y objetiva para poder detectar fraude en los estados financieros o manipulación de datos.

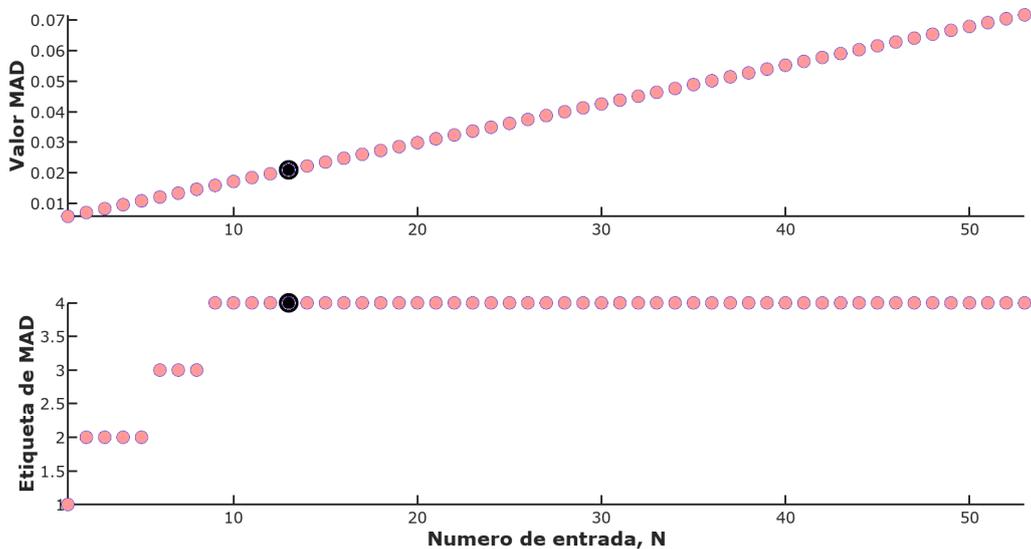


Figura 4.7: Sustitución de los primeros dígitos, 1 por 9, de los primeros 175 valores de la sucesión de Fibonacci. Se resalta el punto donde el valor es una bandera roja.

---

## Homologación de índices estadísticos

A continuación, expondremos como los valores de los índices estadísticos son estandarizados con respecto a la sucesión de Fibonacci y la ley de Benford. La Tabla 4.3 muestra en la primera columna el nombre de los índices estadísticos; luego, en la segunda columna, los valores correspondientes para  $f_{152}$ , que son valores que permiten la afirmación: «los datos se ajustan a la ley de Benford de modo aceptable». La tercera columna presenta los valores de  $f_{152}$  después de la sustitución de los 1s por 9s; para este caso, es claro que la afirmación es: «los datos se ajustan a la ley de Benford de modo no-aceptable». En otras palabras, la tercera columna indica el umbral numérico. Por último, en la cuarta columna indica si el valor umbral debe ser menor, mayor o si se trata de un intervalo. Entonces, al analizar una base de datos cuestionada, se obtendrían los índices estadísticos y tales números se compararían con la Tabla 4.3. Con esta información, el investigador forense contaría con elementos objetivos para indagar más en el proceso electoral.

Es importante señalar que los criterios propuestos en los índices son más laxos que otras alternativas. En efecto, MAD es la base, pero  $R^2$  no es un valor tan cercano a 1, al menos en algunos campos de investigación parece una correlación pequeña. Lo que queda de manifiesto en el valor de la ordenada al origen de la recta ajustada:  $b$  es pequeño, pero no minúsculo. Pero más notorio es la pendiente de la recta ajustada:  $m$  cuenta con un intervalo de aproximadamente 30 %– de tolerancia. Lo interesante de este trió de índices es que están relacionados por construcción, –independientemente del uso de la sucesión de Fibonacci. Por ejemplo, si  $R^2$  es más pequeño que el umbral invalida automáticamente  $b$  y  $m$ . Por su parte, si  $R^2$  es aceptable,  $b$  y  $m$  permiten profundizar en el análisis y junto con la gráfica, es posible conjeturar el tipo de error o manipulación que presenta la base de datos.

Adicionalmente, en este momento, la propuesta se hace para estudiar el primer dígito, pero es fácil de extender para analizar el quinto dígito, o los dos primeros dígitos u otros. Es posible, que el consenso sea que la alteración de la base de datos sea diferente. O bien, que se emplee otra sucesión matemática. En todo caso, la base de datos alterada estándar debe ser más objetiva que un caso empírico, pues puede estar sesgada de algún modo.

Tabla 4.3: Índices estadísticos de los datos Fibonacci-152 y Fibonacci-175 sustituido. Se muestra el criterio de alarma para Fibonacci-175 sustituido

| Indicador | Fibo <sub>152</sub> | Fibo <sub>175sust</sub> | ¿Qué enciende la alarma?            |
|-----------|---------------------|-------------------------|-------------------------------------|
| MAD       | 0.00754             | 0.01687                 | Un valor más grande                 |
| $\chi^2$  | 0.39702             | 2.7174                  | Un valor más grande                 |
| $p_1$     | 0.5287              | 0.004172                | Un valor más pequeño                |
| $R^2$     | 0.9885              | 0.9275                  | Un valor más pequeño                |
| b         | 0.0039              | 0.0284                  | Un valor más grande                 |
| m         | 0.9561              | 0.7383                  | Un valor fuera de: [0.7383, 1.2617] |

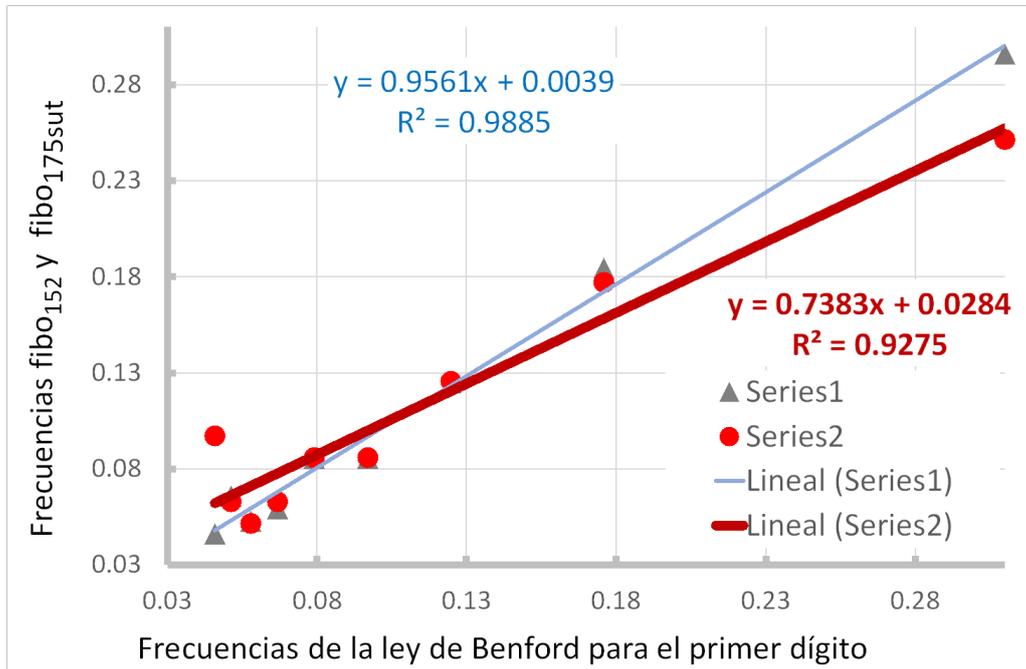


Figura 4.8: Gráficas de regresiones lineales de los datos de la sucesión de Fibonacci para 152 elementos y 175 elementos con intercambio de 1s por 9s.

---

## 4.5. Conclusiones del capítulo

Existen varios criterios para poder detectar posibles fraudes en bases de datos contables. Sin embargo, ¿Cuál es criterio más objetivo para decir si una base de datos está manipulada o no? En este capítulo propusimos un criterio estandarizado con base en la sucesión de Fibonacci, pues cuenta con una alta correlación con ley de Benford y presenta la ventaja de ser una base para modelar diversos fenómenos naturales y sociales. Mostramos que la propuesta permite homologar los índices estadísticos, con el fin de que la interpretación para un abogado o profesional a fin sea más inteligible en la toma decisiones jurídicas. En el siguiente capítulo, mostramos como la propuesta puede ser útil en el análisis e interpretación de bases de datos de tres elecciones presidenciales, dos en México y una en España.

# Capítulo 5

## Análisis electorales del primer dígito

En este capítulo presentamos los resultados del análisis de las bases de resultados electorales (de cada partido) en las elecciones presidenciales de México 2006, España 2015 y México 2018. Para tal fin, estudiamos la distribución del primer dígito de cada elección y obtenemos los índices estadísticos: MAD, prueba-Z para diferencia de proporciones, prueba  $\chi^2$  y linealización ( $R^2$ ,  $m$  y  $b$ ). Después, como criterio de evaluación utilizamos la Tabla 4.3 para identificar en que caso no se supera el umbral límite, lo que implicaría que se requiere un estudio más profundo. Por último, para cada elección realizamos una breve discusión del significado de los resultados.

### 5.1. Elecciones presidenciales de México 2006

Primero, revisamos el caso de las elecciones presidenciales mexicanas del año 2006. La Tabla 5.1 muestra los valores de los siguientes indicadores: Desviación Media Absoluta (MAD),  $\chi^2$ , el valor  $p$  de la prueba-Z de diferencia de proporciones, el coeficiente de correlación lineal de Pearson, la ordenada al origen  $b$  y la pendiente de la recta  $m$  de la recta de regresión lineal. Estos indicadores se calcularon para cada uno de los partidos que participaron en la contienda electoral, en este caso fueron PAN, Por el Bien de Todos, Alianza, Nueva Alianza, Alternativa, además de los votos no-registrados y los votos nulos. A través de la comparación de los valores obtenidos en la Tabla 4.3, se determina si las elecciones pueden mostrar algún tipo de irregularidad que requiera una investigación a mayor profundidades o bien que no se cuentan con elementos de que existió algún tipo de manipulación en la base de datos.

## Análisis MAD para elecciones presidenciales México 2006

Al calcularse el valor de MAD observamos que para todos los partidos, además de los votos no-registrados y votos nulos, los índices calculados superan los umbrales de los valores establecidos en la Tabla 4.3. (calculado mediante  $Fibo_{175}$  sustituido). Lo cual indica que los datos obtenidos se encuentran con mucha variabilidad con respecto a la distribución teórica; es decir, la ley de Benford. De modo que el criterio propuesto por Nigrini tampoco se cumple, pues nuestra propuesta es más laxa.

Tabla 5.1: Valores de las linealizaciones y MAD de los datos electorales de los partidos que participaron en la contienda presidencial de México 2006. En tipo negritas están los indicadores que no cumplieron el criterio de umbral propuesto.

| Indicador | PAN              | POR EL BIEN<br>DE TODOS          | ALIANZA                         | NUEVA<br>ALIANZA |
|-----------|------------------|----------------------------------|---------------------------------|------------------|
| MAD       | <b>0.041307</b>  | <b>0.026137</b>                  | <b>0.052016</b>                 | <b>0.026137</b>  |
| $\chi^2$  | 0.3659           | 0.8667                           | 0.1267                          | 0.8286           |
| $p_1$     | <b>0.0005051</b> | <b>0.008367</b> $\times 10^{-6}$ | <b>0.00157</b> $\times 10^{-6}$ | <b>0.1052</b>    |
| $R^2$     | <b>0.8521</b>    | <b>0.901</b>                     | <b>0.458</b>                    | <b>0.8921</b>    |
| b         | <b>-0.0515</b>   | -0.0227                          | <b>0.0755</b>                   | 0.0032           |
| m         | <b>1.4503</b>    | 1.1941                           | <b>0.3174</b>                   | 1.0193           |
| # Datos   | 15,000,284       | 14,756,350                       | 9,301,441                       | 401,804          |

| Indicador | ALTERNATIVA                      | NO<br>REGISTRADOS | NULOS<br>2006   |
|-----------|----------------------------------|-------------------|-----------------|
| MAD       | <b>0.033476</b>                  | <b>0.022545</b>   | <b>0.051914</b> |
| $\chi^2$  | 0.6918                           | 0.811             | 0.6991          |
| $p_1$     | <b>0.004707</b> $\times 10^{-6}$ | <b>0.7543</b>     | <b>0.01584</b>  |
| $R^2$     | <b>0.8983</b>                    | 0.9342            | <b>0.8252</b>   |
| b         | <b>0.0499</b>                    | 0.015             | <b>0.0419</b>   |
| m         | <b>1.4361</b>                    | 0.8983            | 0.9342          |
| # Datos   | 1,128,850                        | 281,116           | 827,206         |

---

## Análisis con la prueba $\chi^2$ para elecciones presidenciales México 2006

Pretendo determinar a través de una prueba de  $\chi^2$  la existencia de la independencia entre las variables de la base de datos electoral 2006, a saber: los valores de probabilidad para cada dígito del 1 al 9 según el modelo de Benford y los valores relativos correspondientes del primer dígito para cada partido político, así como en los votos nulos.

A partir de la base de datos electoral, descargada del portal del Instituto Nacional Electoral INE, se realizó en el programa *Rstudio* (versión 1.1.456) una prueba de  $\chi^2$  a través de la importación de dicha base (en archivo Excel) al programa y la creación de nueve columnas en una matriz de Excel a las cuales les fue aplicada la función *chisq.test* del programa para obtener los valores de  $\chi^2$  y *p-value* correspondientes al análisis de independencia o no independencia de los valores de probabilidad de Benford y los reportados en cada partido. Es necesario mencionar que como parte de la metodología y para efectos de que la consola de *Rstudio* reconociera la información que se le proporcionó, la probabilidad se expresó en términos de porcentaje y no en valores del 0 al 1.

Asimismo, se debe mencionar que el planteamiento de las hipótesis mediante el que se comprobaron durante el desarrollo de la prueba y que, a través de la observación del *p-value* proporcionado por la función *chisq.test*, permiten establecer conclusiones con una confianza del 99 %.

La hipótesis nula afirma que las variables (antes mencionadas en el planteamiento del problema) mantienen una relación de INDEPENDENCIA, mientras que la hipótesis alterna sugiere que existe entre las variables una relación de NO INDEPENDENCIA.

Para responder al planteamiento del problema con una confianza del 99 % se observa el *p-value* en cada uno de los ocho análisis efectuados, y al superar en todos el valor de 0.01 se concluye que: la relación existente entre los valores de probabilidad de Benford para los dígitos del 1 al 9 y los valores de probabilidad para los dígitos del 1 al 9 de cada una de las nueve variables (PAN, Alianza, Por el bien de todos, Nueva Alianza, Alternativa, no-registrados, nulos) es de INDEPENDENCIA.

Posteriormente, se realizó el cálculo de  $\chi^2$  observándose que los datos obtenidos no superan los valores obtenidos en la Tabla 4.3. Para el caso de los partidos PAN y Alianza, el valor de la  $\chi^2$  es de 0.3659 y 0.1267, respectivamente. Por lo tanto, los valores se encuentran incluso por debajo de los establecidos para *Fibo*<sub>152</sub> y *Fibo*<sub>175</sub> sustituido. En el caso del Partido Por el Bien de Todos y Nueva Alianza, los valores rebasan los valores de *Fibo*<sub>152</sub>, pero no de *Fibo*<sub>175</sub> sustituido, y se consideran que los valores se

---

encuentren dentro de un valor aceptable. Entonces, podemos concluir que no hay discrepancia entre los datos observados y la distribución teórica esperada.

### **Análisis con la prueba-Z de diferencia de proporciones de las elecciones presidenciales 2006**

Para ésta base se aplico la prueba-Z para diferencia de proporciones, con el objetivo de establecer con una confianza del 99%, si existe una diferencia entre las frecuencias reportadas en la Ley de Benford y las frecuencias de cada partido político. Para ello, se utilizó la función *prop.test* en el software *Rstudio* (versión 1.2.5019), creando mediante concatenación de valores, objetos que contienen las frecuencia de Benford y la frecuencia de cada dígito en cada partido.

La teoría estadística señala que, después de desarrollar esta prueba, debe observarse el *p-value* obtenido para aceptar o rechazar las hipótesis planteadas. Cuando el valor-p es mayor que 0.001 se acepta la hipótesis nula, la hipótesis nula ( $H_0$ ) sugiere igualdad entre las frecuencias, mientras que la hipótesis alterna ( $H_a$ ) sugiere desigualdad.

Dicho de otra manera, podemos escribir:

$H_0$ : Frecuencia Benford = Frecuencia partido político.

$H_a$ : Frecuencia Benford distinto a Frecuencia partido político.

Al calcularse el valor-p de la prueba-Z de diferencia de proporciones podemos, observar que para todos los partidos los valores se encuentran muy por debajo del parámetro establecido, es este caso de  $Fibo_{175}$  sustituido, esto nos muestra que la hipótesis nula es rechazada y se acepta como cierta la hipótesis alternativa; es decir, que la distribución teórica y esperada no se ajusta a la ley de Benford.

### **Análisis de regresión lineal para elecciones presidenciales 2006**

En seguida, obtenemos el coeficiente de correlación lineal y sus parámetros de ajuste a una línea recta. En todos los casos, excepto para no-registrados, los valores de correlación son demasiado bajos (utilizando el criterio de la Tabla 4.3). Si el valor de correlación es demasiado bajo, implica que carece de sentido la obtención de la pendiente y la ordenada al origen de la recta ajustada a los datos. De hecho, la Fig. 5.1 muestra (para los dos candidatos con más votos) que las respectivas distribuciones de datos empíricos y el modelo teórico esta lejos de seguir una línea recta.

Empero, comentaremos los demás de los resultados de la linealización. Sobre la ordenada al origen, podemos observar que Por el Bien de Todos, Nueva Alianza y no-registrados son los únicos casos que superan la prueba.

Todos los demás casos muestran valores inaceptables de acuerdo al criterio de umbral establecido en este trabajo.

En lo que respecta a la pendiente de la recta ajustada. Observamos que Por el Bien de todos, Nueva Alianza, no-registrados y Nulos pasan la prueba. El resto de los casos la fallan.

En suma, estos resultados son coherentes con la idea colectiva de que las elecciones presidenciales de México 2006 requerían una investigación más profunda, tal como se expuso en la sección 2.3.1. En términos de investigación, resulta útil encontrar una elección que no siga la ley de Benford y que sea considerada como sospechosa de fraude electoral por alterar sus resultados en la base de datos, pues sirve para poner a prueba este trabajo y futuras metodologías de análisis de bases de datos de otras elecciones.

## 5.2. Elecciones España 2015

A continuación, mostraremos el análisis por MAD, prueba  $\chi^2$ , prueba-Z de diferencia de proporciones, y análisis de regresión lineal de la contienda presidencial española del año 2015.

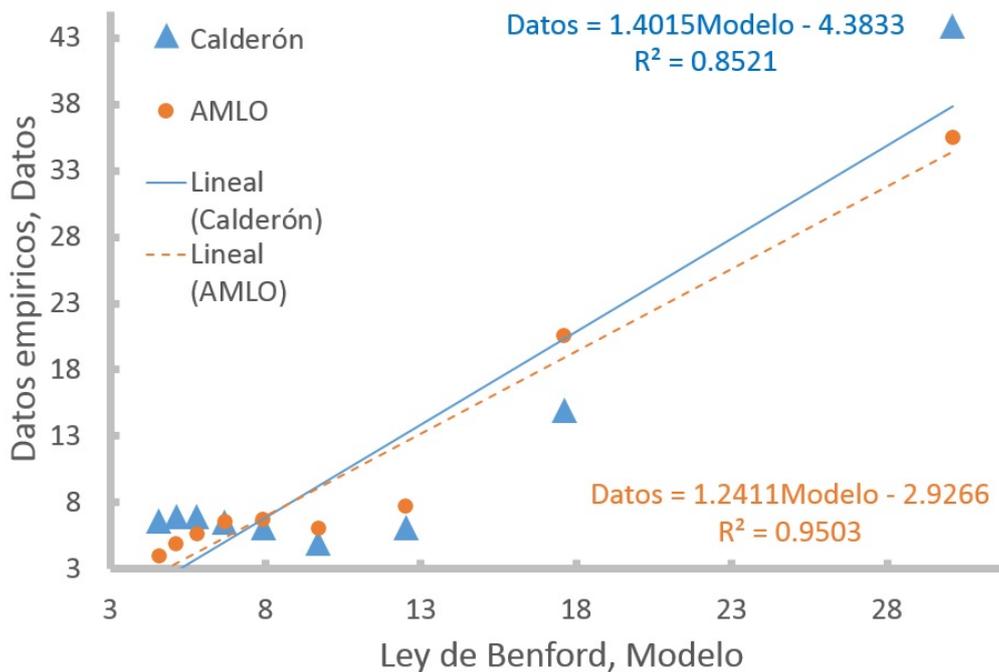


Figura 5.1: Linealización de los datos de los dos candidatos con más votos en la elección presidencial México 2006.

Cómo lo hicimos para el caso antes presentado (México 2006) en la Tabla 5.2 se muestran los cuatro conjuntos de índices estadísticos de trabajo. Es claro solo para el caso donde se cuenta con menos votos, algunos resultados no cumplen con los criterios acordados en la Tabla 4.3. De hecho, puede tratarse de un caso donde observemos efectos de multiplicidad estadística, como se menciona en la sección 3.4. De tal modo, que podemos afirmar que existe congruencia con los datos electorales españoles y la ley de Benford. Y en consecuencia, tenemos una etiqueta de elección limpia. En todo caso mencionaremos algunos detalles sobre los resultados.

### Análisis por MAD de la contienda española

Para el indicador MAD de la contienda española se observa en general que no existe mucha variabilidad ya que no supera los umbrales descrito en la Tabla 4.3

### Análisis por prueba $\chi^2$ de la contienda española

Para el indicador de la prueba  $\chi^2$  podemos notar que ningún partido político supera los umbrales descritos en la Tabla 4.3

### Análisis por prueba-Z para la diferencia de proporciones de la contienda española

Observamos que para la prueba-Z de diferencia de proporciones de la contienda española que todos los casos cumplen con la condición de umbral ya establecida en esta investigación, con excepción de PACMA.

Tabla 5.2: Valores de índices estadísticos de los datos electorales de los partidos que participaron en la contienda de España 2015.

| Indicador | PP        | PSOE      | EQUO      | C         | PACMA          |
|-----------|-----------|-----------|-----------|-----------|----------------|
| MAD       | 0.0056188 | 0.0037398 | 0.001744  | 0.002429  | 0.011451       |
| $\chi^2$  | 0.16187   | 0.091108  | 0.013685  | 0.045679  | 2.2047         |
| $p_1$     | 0.4023    | 0.7614    | 0.9421    | 0.9698    | <b>0.00147</b> |
| $R^2$     | 0.9939    | 0.9958    | 0.9996    | 0.9986    | 0.9983         |
| b         | 0.0065    | -0.0003   | 0.000030  | 0.0002    | <b>-0.0375</b> |
| m         | 0.9333    | 0.994     | 0.9909    | 0.9932    | <b>1.3257</b>  |
| # Datos   | 7,941,236 | 5,443,846 | 3,227,123 | 3,141,570 | 853,102        |

## Análisis de regresión lineal de la contienda española

Como se puede ver la Fig. 5.2 se muestran los datos de la elección presidencial de España 2015 linealizados con la ley de Benford aquí se toman cinco partidos políticos españoles el PP, PSOE, ECUO, C Y PACMA. Se puede observar que los 5 partidos tienen una correlación muy cercana 1 puesto que PP tiene una correlación de 0.9939, PSOE tiene una correlación de 0.9958, EQUO tiene una correlación de 0.9996, C tiene una correlación de 0.9986 y por último PACMA tiene una correlación de 0.9983. Otro indicador interesante son los valores de la ordenada al origen  $b$  y el valor de la pendiente  $m$ . todos los partidos políticos muestran pendientes cercanas a 1 y las ordenadas al origen que se aproximan a 0 concuerda con lo descrito el capítulo cuatro. Con excepción de PACMA, que cumple con la correlación, pero no con los valores de  $b$  y  $m$ .

Respecto al Partido PACMA, no cumple con los criterios en los siguientes indicadores: valor del  $p$ -value, el valor de  $b$  y el de la pendiente. El valor de  $p$ -value o  $p_1$  es de 0.00147, la diferencia con respecto a  $fib_{175}$  sustituido es

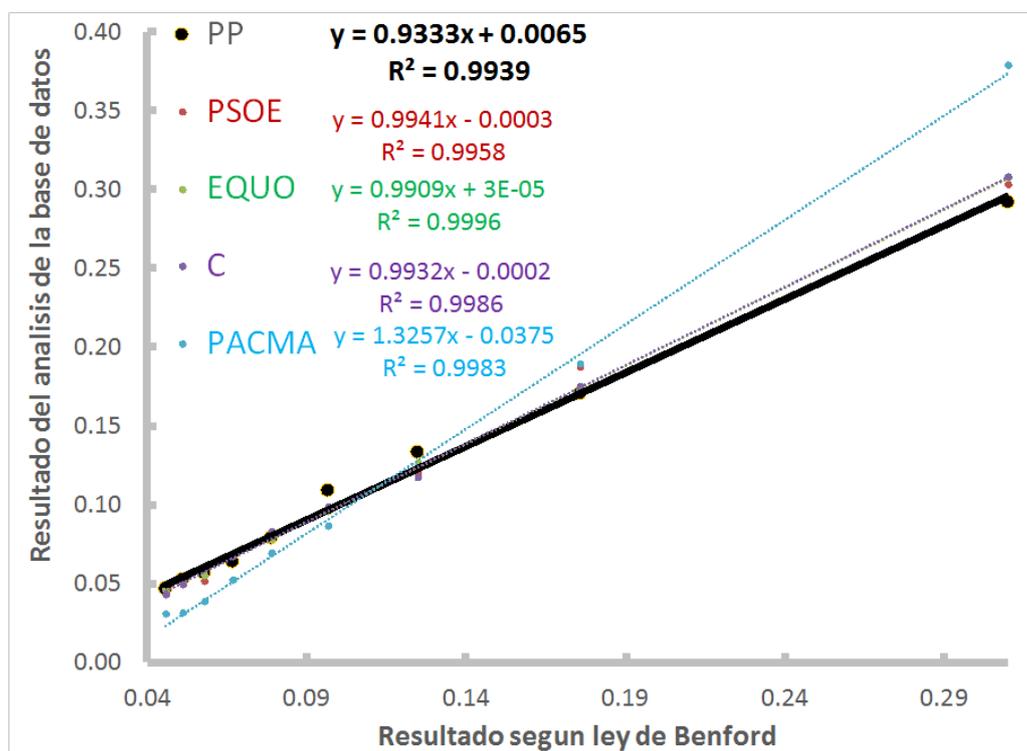


Figura 5.2: Datos de la elección presidencial de España 2015 linealizados con la ley de Benford.

---

de 0.002702, un valor considerablemente bajo. Respecto al valor de  $b$ , este es de -0.0375, se encuentra fuera del umbral. Y en lo que atañe al valor de la pendiente este supera el rango con una diferencia de 0.064. Por lo tanto, se puede decidir investigar más a profundidad la base de datos de este partido o buscar más información para resolver porque no cumplió con los criterios estadísticos.

Con todo, podemos calificar de limpia la elección presidencial de España 2015. Los tres criterios que no cumple PACMA parecen ser un efecto de multiplicidad estadística. Además, carecen de significancia para el resultado neto presidencial. Este resultado global es coherente con la literatura especializada [5]. Así, contamos con una elección que sigue la ley de Benford y es etiquetada como limpia por la sociedad. Para esta y futuras investigaciones puede utilizarse como una referencia para indagar más a profundidad en otras bases de datos y algoritmos de estudio. Por último, la base de datos se puede consultar en el sitio oficial de las elecciones de España 2015.

### 5.3. Elecciones presidenciales México 2018

A continuación, presento el análisis de la contienda presidencial México 2018 en la cual participaron MORENA, PAN, PRI, PRD, PVEM, PT, MC, PANAL, y PES; además presentamos los resultados de los votos nulos. Mostraré el análisis por MAD, prueba  $\chi^2$ , prueba-Z para diferencia de proporciones, y análisis por regresión lineal.

Utilizaremos la base de datos presidencial 2018.xlsx. en cual tiene información compactada y sintetizada de los partidos políticos descargada del Instituto Nacional Electoral y sus respectivos cálculos de frecuencias del primer dígito. En éstos cálculos se muestran las frecuencias de cada dígito del uno al nueve, comparada con la ley de Benford y los resultados de los partidos políticos todas ellas obtenidas con el software Octave.

#### **Análisis por MAD de las elecciones presidenciales mexicanas 2018**

Observaos que que cada partido de la contienda supero el umbral MAD establecido en la Tabla 4.3. En este punto se podría inferir una sospecha de manipulación en la base de datos electoral.

#### **Análisis por prueba $\chi^2$ para la contienda la contienda elecciones mexicanas 2018**

Para la prueba  $\chi^2$  todos los partidos encendieron las alarmas excepto en el PAN con un valor de 2.425 y PRD con un valor de 1.7353.

## Análisis por prueba-Z de diferencia de proporciones para la contienda elecciones mexicanas 2018

Para ésta base nuevamente como en las elecciones 2006 se aplicó la prueba-Z para diferencia de proporciones, con el objetivo de establecer con una confianza del 99 %, si existe una diferencia entre las frecuencias reportadas en la Ley de Benford y las frecuencias de cada partido político. Para ello, se utilizó la función *prop.test* en el software *Rstudio* versión 1.2.5019, creando mediante concatenación de valores, objetos que contienen las frecuencia de Benford y la frecuencia de cada dígito en cada partido.

Cómo vimos en las elecciones pasadas del año 2006. La teoría estadística señala que, después de desarrollar esta prueba, debe observarse el *p-value* para aceptar o rechazar las hipótesis planteadas. Cuando el valor-p es mayor que

Tabla 5.3: Valores de las linealizaciones de los datos electorales de los partidos que participaron en la contienda presidencial de México 2018.

| Indicador | MORENA                                  | PAN             | PRI                                     | PRD                                     | PVEM            |
|-----------|---|-----------------|---|---|-----------------|
| MAD       | <b>0.279901</b>                         | <b>0.104055</b> | <b>0.201817</b>                         | <b>0.069319</b>                         | <b>0.078107</b> |
| $\chi^2$  | <b>17.701</b>                           | 2.425           | <b>8.6865</b>                           | 1.7353                                  | <b>4.4124</b>   |
| $p_1$     | <b><math>1.2 \times 10^{-13}</math></b> | <b>0.001842</b> | <b><math>6.7 \times 10^{-16}</math></b> | <b><math>6.7 \times 10^{-16}</math></b> | 0.3361          |
| $R^2$     | <b>0.907</b>                            | 0.9439          | <b>0.4409</b>                           | 0.9339                                  | <b>0.9073</b>   |
| b         | <b>0.0873</b>                           | <b>0.0358</b>   | <b>0.0594</b>                           | 0.0023                                  | 0.0105          |
| m         | <b>1.7703</b>                           | <b>0.672</b>    | <b>0.4615</b>                           | 0.9704                                  | 1.085           |
| # Datos   | 19,796,658                              | 8,036,231       | 6,082,837                               | 1,182,529                               | 729,252         |

| Indicador | PT                                      | MC              | PANAL           | PES                                    | NULOS           |
|-----------|---|-----------------|-----------------|--|-----------------|
| MAD       | <b>0.212783</b>                         | <b>0.126316</b> | <b>0.116361</b> | <b>0.149606</b>                        | <b>0.142063</b> |
| $\chi^2$  | <b>9.2138</b>                           | <b>6.5854</b>   | <b>10.088</b>   | <b>5.7961</b>                          | <b>7.6846</b>   |
| $p_1$     | <b><math>5.6 \times 10^{-11}</math></b> | 0.006153        | 0.08273         | <b><math>2.2 \times 10^{-5}</math></b> | 0.006372        |
| $R^2$     | 0.95                                    | <b>0.8041</b>   | 0.9405          | <b>0.8377</b>                          | <b>0.7615</b>   |
| b         | <b>-0.0713</b>                          | 0.0011          | <b>0.0439</b>   | <b>0.0534</b>                          | 0.0004          |
| m         | <b>1.6276</b>                           | 0.9811          | <b>1.3832</b>   | <b>0.5146</b>                          | 0.9947          |
| # Datos   | 2,243,330                               | 699,861         | 445,951         | 821,714                                | 1,299,974       |

---

0.001 se acepta la hipótesis nula, la hipótesis nula ( $H_0$ ) sugiere igualdad entre las frecuencias, mientras que la hipótesis alterna ( $H_a$ ) sugiere desigualdad.

Dicho de otra manera:

$H_0$ : Frecuencia Benford = Frecuencia partido político

$H_a$ : Frecuencia Benford distinto a Frecuencia partido político

Para éste caso la prueba mostró resultados interesantes. Se puede observar que los partidos como MORENA, PAN, PRI, PRD, PT Y PES muestra umbrales por debajo de lo establecido, mientras que los partidos como PVEM, MC, PANAL Y los registros NULOS mostraron superar los umbrales construidos en la Tabla 4.3.

### **Análisis por regresión lineal de la contienda elecciones mexicanas 2018**

En la Fig. 5.3 se presenta el análisis por linealización entre la ley de Benford y la base de datos de la elección de México 2018, donde sólo se presentan tres partidos con mayor número de votos, ello con el fin de mostrar los datos más significativos<sup>1</sup>.

Podemos observar en ésta gráfica la relación lineal que existe entre la ley de Benford y los resultados del análisis de la base de datos, de los tres partidos políticos más “importantes” PRI, PAN Y MORENA. en las elecciones electorales del pasado año 2018.

Podemos vislumbrar tres correlaciones sobre el comportamiento de la ley de Benford con respecto a la Base de datos estudiada. En el PRI podemos notar que existe una correlación menor de 0.4409, PAN una correlación de 0.9439 Y MORENA con 0.907.

Se puede notar un contraste significativo entre PRI, PAN Y MORENA puesto que PRI tiene una correlación menor que 0.5 mientras que PAN Y MORENA muestran una correlación cercanos A 1 lo que significa que la relación entre la ley de Benford y la base de datos estudiada del PRI no explica el modelo lineal. Sin embargo MORENA Y PAN si explican el modelo.

Ahora bien, la comparación son muy similares con respecto al modelo de regresión lineal  $y = mx + b$  cuando  $m = 1$  y  $b = 0$ , *i.e.* tenemos la función identidad  $y = x$ . En este modelo « $y$ » representaría las bases de datos

---

<sup>1</sup>Carece de sentido ajustar los nueve dígitos y sus frecuencias a un polinomio. Nueve datos de ajuste a una recta pueden ser polémico para algunos académicos. En el quehacer diez datos para ajustar una recta suele ser aceptable. En tal sentido, un ajuste mediante un polinomio de grado dos requeriría veinte datos. Pero en nuestro caso el asunto del ajuste lineal trasciende la convención estadística. Al comparar datos empíricos con el modelo teórico, el mejor modelo de ajuste es el lineal. Pues se espera que la teoría describa la observación real de un modo uno a uno. Si el comportamiento no es lineal, entonces, es posible que se requiera un ajuste en la teoría.

---

estudiadas y  $x$  la ley de Benford. El PRI tiene valores de  $m = 0.4615$  y  $b = 0.0594$ , que es cercano a cero.

El PAN (Partido Acción Nacional) podemos ver que los valores de  $m$  y  $b$  son muy cercanos a 1 y 0 ya que su valor de  $m = 0.672$  y  $b = 0.0358$  y con una alta correlación de 0.9439.

MORENA (Movimiento de Regeneración Nacional ) tiene valores de  $m = 1.7703$  algo lejano a 1 con respecto a PAN y PRI pero con una correlación de 0.907 más baja que la del PAN.

1. Podemos decir que las correlaciones del PAN Y MORENA explican muy bien la concordancia entre la ley de Benford y las bases de datos estudiadas pero la correlación del PRI no explica muy bien el contraste entre dicha ley Benford y las bases de datos estudiadas.
2. MORENA aunque tuvo una correlación cercana a 1 podemos ver que la pendiente esta alejada de 1 de hecho es casi 2 y la ordenada al origen  $b$  está alejada de 0 lo que significa que la relación entre la ley de Benford vista de ser muy cercana
3. EL PAN tiene una correlación mucho más cercana a 1 que MORENA al igual que la ordenada al origen lo que significa que se la base de Datos del PAN se ajusta más a la ley de Benford.

Cabe destacar que en la Tabla 5.3 se pude notar que se encendieron todas las alarmas de casi todos los partidos políticos. Otro criterio interesante de observar con respecto a la Tabla 4.3 es que el valor de correlación del PAN no enciende la alarma pero sus valores de la ordenada al origen  $b$  y pendiente  $m$  si lo hacen, lo cual demuestran que no aproximan al modelo dicha relación. Este tipo de análisis u observación es mas completo ya que a pesar de que exista una buena correlación con ley de Benford con respecto a la correlación lineal los demás parámetros impiden que sea una buena aproximación ya que superan los umbrales de alarma.

Ahora bien es muy notable que al tomar los resultados de todos los partidos políticos casi todas las alarmas se encendieran: Se podría interpretar que algo en el proceso no se realizó correctamente. Sin embargo, más allá de ideologías, esta metodología se usó con resultados coherentes con las elecciones de México 2006 y España 2015.

Por lo tanto, se concluye que las elecciones del 2018 requieren un análisis más profundo para descartar la probabilidad de manipulación en la base de datos.

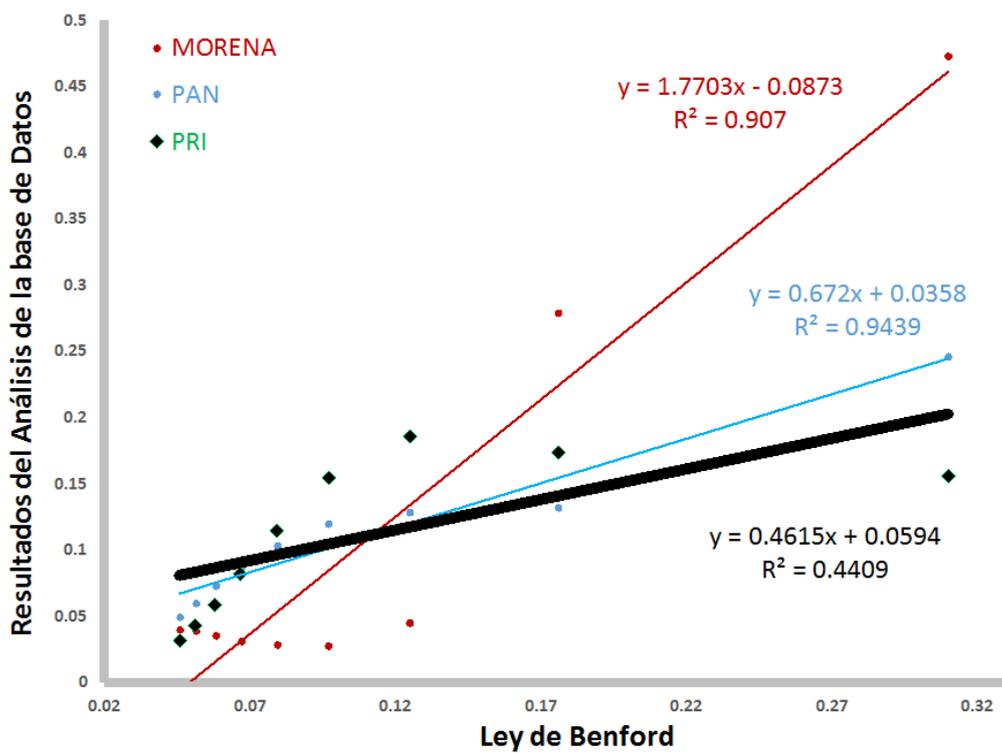


Figura 5.3: Datos de la elección presidencial de México 2018 linealizados con la ley de Benford. Solo se muestran ajustes para tres partidos políticos.

---

## 5.4. Conclusiones de capítulo

Para los casos de México, puede ser que las dos bases de datos estén manipuladas , lo que implica que se necesita una revisión más detallada de ambas bases de dato, años: 2006 y 2018.

Puede ser que las elecciones mexicanas no se ajusten a los criterios estadísticos, posiblemente por variables ocultas. Lo que implica que 2006 y 2018 están limpias.

Puede ser que el criterio estadístico es demasiado rigorista. Por lo que se necesita hacer un estudio de otros fuentes u otras pruebas de bondad de ajuste como el Criterio de Cramér-von Mises.

Para España podemos notar que uno de sus partidos resalta tres índices de posible manipulación. Pero su porcentaje de importancia es pequeño. Si se quiere considerar un estudio global de las elecciones, entonces podemos concluir que este estudio es congruente con estudios previos en el sentido de que la base estudiada se encuentra limpia.

# Capítulo 6

## Conclusiones generales

Con base a la evidencia presentada en este trabajo, podemos listar nuestras principales conclusiones:

- El estudio de bases de datos vía la frecuencia de aparición de dígitos dentro de una posición puede ser importante para dictaminar si la información fue alterada. Por ello se requiere definir umbrales numéricos para que los índices estadísticos sin rodeos se interpreten al momento de una alerta. Lamentablemente, en muchos casos estas pautas provienen de inducciones influenciadas por la subjetividad, e.g. MAD. Por ello, proponemos utilizar la sucesión de Fibonacci-sustituido como un calibrador de los índices estadísticos; logrando con ello: objetividad y claridad metodológica.
- El calibrador con la sucesión de Fibonacci sustituida es efectivo para definir el comportamiento de una base de datos que al inicio tendía a la ley de Benford es alterada para luego ser incongruente a dicha regla. Por medio de tal base de datos es sencillo definir umbrales de alerta para cualquier índice estadístico y su interpretación es sencilla, por lo que es adecuado como medio didáctico para explicarlo para perfiles profesionales con escasos conocimientos en estadística.
- Estudiamos tres bases de datos de elecciones presidenciales. Observando que el criterio propuesto con la sucesión de Fibonacci-sustituido es laxo y coherente respecto a otras observaciones en la literatura. De hecho, nos permitió observar el fenómeno de multiplicidad estadística, Por ejemplo, que para resultados partidarios poco representativos de la elección general se superaba el umbral de alerta. Por ello, podemos decir que este tipo de estudios es útil para guiar cuando se debe realizar una investigación más profunda ante la sospecha de fraude electoral.

## 6.1. Prospectiva de la investigación

Para ello al corto plazo se pueden realizar varias acciones, como las siguientes:

- Si bien este análisis se realizó sólo para el primer dígito, dicho estudio se puede generalizar para el segundo dígito y combinaciones de los dos primeros dígitos. Además observar si existe una distribución en la frecuencia de los últimos dígitos de la base de datos.
- En éste documento se realizó una investigación a nivel nacional de tres elecciones, en dos países. Sin embargo se puede profundizar el trabajo a nivel estatal y municipal con el fin de profundizar en la pesquisa. Esperando con ello que las anomalías sean más evidentes. Pues el resultado general, al ser una mezcla de bases de datos, debe tender más a la ley de Benford en comparación con las bases de datos componentes, como muestra la Fig 1.6. Por tanto para el desarrollo de esta propuesta se requieren estudiar más bases de datos.
- Sería interesante estudiar bases de datos sintéticas construidas de modos diversos y alteradas a diferentes niveles. Ello con el fin de poner a prueba doble ciego la propuesta del calibrador de Fibonacci-sustituido y la ley de Benford como medios para identificar posibles adulteraciones de los datos.

## 6.2. Palabras finales

La ley de Benford es una herramienta matemática muy útil para detectar anomalías que pudieran relacionarse con fraude, pero se requiere de un buen calibrador para definir umbrales de alerta. Aún así al momento de interpretar los resultados con la ayuda de métodos estadísticos se debe mantener la objetividad de su significado cuando no se sigue la ley de Benford: la existencia de una posible manipulación en la base de datos.

Este documento pretende realizar un estudio con dichas características para que un operador forense o abogado pueda usarlo para futuros dictámenes, con el objetivo de ser más claro y así obtener una interpretación más rígida pero inteligible con respecto al fundamento del estudio.

# Bibliografía

- [1] Boyle. *An application of Fourier series to the most significant digit problem*. 9.<sup>a</sup> ed. American Mathematical Monthly, 1994.
- [2] Gonzalo Castañeda. «La ley de Benford y su aplicabilidad en el análisis forense de resultados electorales». En: *Política y gobierno* 18.2 (2011), págs. 297-329.
- [3] Lucas Lacasa Juan Fernández-Gracia. *Bipartisanship breakdown, functional networks and forensic analysis in Spanish 2015 and 2016 national elections*. Dic. de 2017.
- [4] Boudewijn F. Roukema. *A first-digit anomaly in the 2009 Iranian presidential election*. Mayo de 2018.
- [5] Francisco R. Villatoro. *Análisis forense de las elecciones españolas del 20-D y del 26-J*. Mayo de 2018.
- [6] Jr. Walter R. Mebane. *Political Science 485 (Fall 2017)*, “Election Forensics”. Mayo de 2018.
- [7] Emiliano Calderón Luis Horacio Gutiérrez. *La ley de Benford para el segundo dígito y los resultados electorales en México*. Mayo de 2018.
- [8] María Caputi Zunini. *VI Coloquio de Matemática*. Dic. de 2017.
- [9] Grupo Cynthus. *Idea y el uso de la ley de Benford*. Nov. de 2017.
- [10] María del Socorro Brito. *Identificación de elementos de auditoría*. Mayo de 2018.
- [11] Emerging Technology from the arXiv. *Benford’s Law And A Theory of Everything*. Mayo de 2010.
- [12] A. Jackson M. Sambridge H. Tkalčić. *Benford’s law in the natural sciences*. Nov. de 2010.
- [13] Bo-Qiang Ma Lijing Shao. *The Significant Digit Law in Statistical Physics*. Mayo de 2010.

- [14] Carlo Altamirano y Alberto Robledo. *Possible thermodynamic structure underlying the laws of Zipf and Benford*. Mayo de 2018.
- [15] Aisha Dantuluri<sup>1</sup> y Shantanu Desai. *Do tau Lepton Branching fractions obey Benford's Law?* Mayo de 2018.
- [16] David DiBattista Aaron D. Slepko Kevin B. Ironside. *Benford's Law: Textbook Exercises and Multiple-choice Testbanks*. Nov. de 2013.
- [17] Tariq Ahmad Mir. *Citations to articles citing Benford's law: a Benford analysis*. Mar. de 2016.
- [18] Carlos Andrés Altamirano Allende. *Leyes de potencias bajo la mecánica estadística no extensiva : Ley de Zipf y Ley de Benford*. Mayo de 2018.
- [19] Franklin Geovanny Toalombo Montero. *Análisis de los resultados electorales obtenidos de la aplicación del voto electrónico en el Ecuador, caso proceso electoral de febrero de 2014*. 2016.
- [20] Franklin Geovanny Toalombo Montero. *Análisis de los resultados electorales obtenidos de la aplicación del voto electrónico en el Ecuador, caso proceso electoral de febrero de 2014*. Trabajo de investigación como requisito previo a la obtención del título de Magister en Gestión Pública. Quito: IAEN. 117 p., 2016.
- [21] Steven J. Miller. *Benford's law : theory and applications*. 1.<sup>a</sup> ed. Princeton University Press., 2015. ISBN: 9780691147611.
- [22] T. P. Hill. *The first-digit phenomenon*. 86.<sup>a</sup> ed. Am. Sci., 1998.
- [23] Carlo Altamirano y Alberto Robledo. *Possible thermodynamic structure underlying the laws of Zipf and Benford*. Mayo de 2018.
- [24] et. al. Ed. Bormashenko E. Shulzinger. *Benford's Law, its applicability and breakdown in the IR Spectra of polymers*. Mayo de 2018.
- [25] Wei Su Dongdong Fua Yun Q. Shia. *A generalized Benford's law for JPEG coefficients and its applications in image forensics*. Jun. de 2018.
- [26] Anthony TS Ho Yun Q Shi Xi Zhao. *Image forensics using generalised benford's law for accurate detection of unknown jpeg compression in watermarked images*. Jun. de 2018.
- [27] César Carrera. *La Ley de Benford y su aplicación a la economía*. Ago. de 2018.
- [28] H.R. Varian. «Benford's Law.» En: *The American Statistician* 26 (1972), págs. 65-66.

- [29] Stefan Gunnel y Karl-Heinz Tödter. «Does Benford's Law Hold in Economic Research and Forecasting?» En: *Discussion Paper Series 1: Economic Studies* 2007,32 (2007), págs. 453-488.
- [30] Statements on Auditing Standards. *Consideration of Fraud in a Financial Statement Audit*. Ago. de 2018.
- [31] Statements on Auditing Standards. *Analytical Procedures*. Ago. de 2018.
- [32] R. Mansilla. *Análisis de los resultados electorales a partir de la Ley de Benford*. 2006.
- [33] Zoraida Garcia. *Ciencia forense en el contexto de nuevo sistema de judicial penal*. 1.<sup>a</sup> ed. Tribunal Superior de Justicia de la Ciudad de México, Red temática de Ciencias Forenses., 2016. ISBN: 978-607-9082-20-8.
- [34] Comisión de Venecia. *Sobre la identificación de irregularidades electorales por medio de métodos estadísticos redactado por la Comisión de Venecia para la Comisión europea por la democracia del año 2018*. 2018.
- [35] Neus Canal Díaz. *Comparación de proporciones*. 2020.
- [36] Francisco Sánchez Villarreal. *Estadística 2 Análisis de regresión simple*. 2012. URL: Facultad%20de%20Ciencias,%20UNAM.
- [37] Luis Rosero Armijos Fernando Morena Brieva Nelson Cordova Rosas. *Fibonacci en los Negocios .Acercas de un modelo matemático para pronósticos financieros "*. 2020.

# Anexos

# Apéndice A

## Apéndice: Códigos octave

A continuación, a modo de ser transparentes con los métodos utilizados en este trabajo, y fortalecer el compartir la información, presentamos los códigos esenciales de los programas realizados. Debemos mencionar, que si bien realizamos una amplia documentación en los *scripts*, se omitieron los acentos en las palabras.

### A.1. Disección de los códigos

#### A.1.1. Programa de la moneda

```
1 clc; clear; close all %limpieza de memoria
2 tic %inicia un reloj
3 x = rand(1, 2000); %generacion de numeros aleatorios
4 %simulacion de probabilidad en moneda justa
5 cara = x < 0.5; %Una cara de la moneda
6 cruz = x > 0.5; %La otra cara de la moneda
7 ganancia = cumsum(cara - cruz); %Considerando que se gana una
   moneda, se pierde una moneda
8 plot(ganancia);
9
10 S = zeros(size(ganancia));
11 for i=2:length(ganancia)
12     if (cara(i-1) == cara(i));
13         S(i) = 1 + S(i-1);
14     else
15         S(i) = 0;
16     end
17 end
18 fin = [cara; S];
19 testBenfordConformity(S)
20
```

```

21 %http://math.boisestate.edu/~wright/courses/matlab/lab_13/html/
    lab_13.html
22
23 %% contador-vectorizado de rachas
24
25 %Sall = cumsum(a);
26 %stupidx = find(diff(a)==-1)+1;
27 %a2=a;
28 %a2(stupidx) = -Sall(stupidx)+[0 Sall(stupidx(1:end-1))];
29 %S2 = cumsum(a2);
30 %S = [0 S2(1:end-1)];
31 % https://stackoverflow.com/questions/15628887/efficient-ways-to
    -count-a-streak-of-consecutive-integers-in-matlab
32
33 %a00*50/(sum(a)/2) % error porcentual
34
35 toc %termina el reloj

```

### A.1.2. Los programas adoptados por Marco Cococcioni

```

1 function demo
2
3 %%PART 1
4 disp('Plotting the Benford''s Law (see at the last figure)');
5 plotBenfordLaw
6 disp('Press any key to continue...');
7 pause
8
9 disp('computing the first digit for the array [5.23, 0.00035,
    -83.24, 0.0000064, 1936, 123]');
10 disp('(it is expected to display [5 3 8 6 1 1] )');
11 firstDigit = calcFirstDigit([5.23, 0.00035, -83.24, 0.0000064,
    1936, 123]);
12 disp(firstDigit);
13 disp('Press any key to continue...');
14 pause
15
16 %%PART 2
17 disp('Testing the conformity to the Benford''s Law for three
    kinds of distributions');
18 disp('');
19 disp('First distribution: LogNormal (it MUST conform!)');
20 testBenfordConformity(lognrnd(1,2,1e5,1));
21 title('Log-Normally distributed pseudo-random numbers CLOSELY
    CONFORM to Benford''s Law');
22 disp('Press any key to continue...');
23 pause
24
25 disp('Second distribution: Uniform (it DOES NOT CONFORM to

```

```

    Benford''s Law)');
26 testBenfordConformity(rand(1e5,1));
27 title('Uniformly distributed pseudo-random numbers DO NOT CONFORM
    to Benford''s Law');
28 disp('Press any key to continue...');
29
30 disp('Third distribution: Normal (it DOES NOT CONFORM, but
    coarsely resemble Benford''s Law)');
31 testBenfordConformity(randn(1e5,1));
32 title('Normally distributed pseudo-random numbers COARSELY
    RESEMBLE to Benford''s Law');
33
34 disp('');
35 disp('(end of the demo)');

1 function [fistDigit , normNumOfOccOfEachDigit] = calcFirstDigit(x,
    plotTheNormalizedNumberOfOccurrences)
2 % calcFirstDigit - computes the first significant digit of the
    vector x given as input
3 %
4 %% Example 1
5 % fistDigit = calcFirstDigit([5.23, 0.00035, -83.24, 1936, 123],
    true);
6 % disp('Should display [5 3 8 1 1]');
7 % disp(fistDigit');
8 %
9 %% Example 2
10 % [~, normNumOfOccOfEachDigit] = calcFirstDigit([5.23, 0.00035,
    -83.24, 1936, 123]);
11 % disp('Should display [.4 0 .2 0 .2 0 0 .2 0]');
12 % disp(normNumOfOccOfEachDigit');
13 %
14 % See also HISTC
15
16 % Copyright Marco Cococcioni, 2015
17
18 if nargin < 2,
19     plotTheNormalizedNumberOfOccurrences = false;
20 end
21
22 %convert the input into a double vector
23 x = double(x(:));
24
25 ax = abs(x(x~=0 & not(isinf(x)) & not(isnan(x))));
26 %per prendere solo i valori diversi da 0 e da Inf e in valore
    assoluto
27 fistDigit = floor(ax./(10.^floor(log10(ax))));
28
29 if ( nargin > 1 ) || plotTheNormalizedNumberOfOccurrences,

```

```

30     normNumOfOccOfEachDigit = histc(fistDigit,1:9);
31     normNumOfOccOfEachDigit = normNumOfOccOfEachDigit./sum(
normNumOfOccOfEachDigit);
32 end
33
34 if plotTheNormalizedNumberOfOccurrences,
35     bar(1:9, normNumOfOccOfEachDigit);
36     title('Normalized number of occurrences for each digit from 1
to 9');
37     xlabel('First Digit (1-9)');
38 end
39
40 end

1 function plotBenfordLaw
2 %% This function plots the Benford's Law First Digit distribution
3 % More info can be found on Wikipedia:
4 % https://en.wikipedia.org/wiki/Benford%27s\_law
5
6 % Copyright Marco Cococcioni, 2015
7
8 figure('Name','Benford's Law');
9 benford_dist = log10(1 + (1./(1:9)))';
10 bar(1:9, benford_dist, 'r');
11 for i=1:9,
12     text(i-.25,benford_dist(i)+0.01, sprintf('%.1f%%, 100*
benford_dist(i)));
13 end
14 title('Benfor's Law (also known as First-Digit Law)');
15 xlabel('First Digit (1-9)');
16 ylabel('Predicted number of occurrences (normalized)');

1 function [degreeOfConformity, MAD] = testBenfordConformity(x)
2 % This function computes the degree of conformity of the given
vector x to
3 % Benford's Law. More precisely, it compares the occurrences of
the first
4 % digits in the vector with the Benford's Law using the Mean
Absolute Deviation
5 % criterion.
6 %
7 % Input:
8 %     x: the vector we want to know if it follows or not the
Benford's Law
9 %
10 % Output:
11 %     degreeOfConformity: the degree of conformity
12 %                       1 = close conformity
13 %                       2 = acceptable conformity

```

```

14 %           3 = marginal conformity
15 %           4 = non conformity
16 %
17 %           MAD:   the Mean Absolute Deviations (MAD) of observed
           occurrences
18 %           with respect to from Benford's first digit
           distribution
19 %
20 %A reference to MAD criterion and the used thresholds is the
           following:
21 %
22 %M Nigrini & JT Wells, Benford's Law: Applications for Forensic
           Accounting,
23 % Auditing, and Fraud Detection, Wiley, 2012 (page 160)
24 %
25 %
26 %% Example 1
27 % testBenfordConformity(lognrnd(1,2,1e5,1));
28 % title('Log-Normal random numbers show a CLOSE CONFORMITY to
           Benford's Law');
29 %
30 %% Example 2
31 % testBenfordConformity(rand(1e5,1));
32 % title('Uniform random numbers DO NOT CONFORM to Benford's Law
           ');
33 %
34 %% Example 3
35 % testBenfordConformity(randn(1e5,1));
36 % title('Normal random number DO NOT CONFORM to Benford's Law (
           even if it coarsely resembles it)');
37 %
38 % See also calcFirstDigit
39
40 % Copyright Marco Cococcioni, 2015
41
42
43 benford_norm_occ = log10(1 + (1./(1:9))); %distribuzione di
           benford
44
45 %Calcolo la frequenza relativa della prima cifra significativa
46
47 [~, given_data_norm_occ] = calcFirstDigit(x);
48
49 % computing the Mean Absolute Deviation
50 MAD = mean(abs(given_data_norm_occ - benford_norm_occ));
51
52 % The thresholds used to establish the degreeOfConformity
53 % depending on the value of MAD
54 madClose = 0.006; % threshold for close conformity

```

```

55 madAcceptable = 0.012; % threshold for acceptable conformity
56 madMarginally = 0.015; % threshold for marginal conformity
57
58
59
60 %% Plot the reference distribution and the given distribution
61 both = [benford_norm_occ, given_data_norm_occ];
62 figure;
63 bar(both);
64 leg=legend('First Digit distribution according to Benford', '
    First digit distribution for the given dataset');
65 set(leg, 'fontsize',16);
66
67 if (MAD <= madClose)
68     xlabel(sprintf('The Mean Absolute Deviation from Benford''
    Law is = %g [CLOSE CONFORMITY]',MAD));
69     degreeOfConformity = 1;
70 elseif (MAD <= madAcceptable)
71     xlabel(sprintf('The Mean Absolute Deviation from Benford''
    Law is = %g [ACCEPTABLE CONFORMITY]',MAD));
72     degreeOfConformity = 2;
73 elseif (MAD <= madMarginally)
74     xlabel(sprintf('The Mean Absolute Deviation from Benford''
    Law is = %g [MARGINAL CONFORMITY]',MAD));
75     degreeOfConformity = 3;
76 else
77     xlabel(sprintf('The Mean Absolute Deviation from Benford''
    Law is = %g [NON CONFORMITY]',MAD));
78     degreeOfConformity = 4;
79 end

```

### A.1.3. Código de sustitución de primeros dígitos y análisis MAD

```

1 clear, clc, close all % limpieza de memoria
2 x = fibo(175); % a 75 umbral del cambio
3 y = calcFirstDigit(x, true);
4 z = find(y==1); % encontrar la posición de primeros 1
5 a = zeros(1, length(z)); % vectorización
6 b = a;
7 % sustitución controlada, los dígitos 1 cambian a dígitos 9
8 for i = 1:length(z)
9     y(z(i)) = 9;
10    [a(i),b(i)] = testBenfordConformity(y);
11    close all % cerrando ventanas del comando anterior
12 end
13 % Ver el proceso de cada sustitución
14 subplot(2,1,1)

```

```
15 plot(b, 'o', 'MarkerSize', 20, 'MarkerFaceColor', [1 .6 .6], '
    MarkerEdgeColor', 'blue')
16 ylabel('Valor MAD', 'FontSize', 30, 'FontWeight', 'bold')
17 set(gca, 'FontSize', 24, 'linewidth', 3) ; axis tight
18
19 subplot(2,1,2)
20 plot(a, 'o', 'MarkerSize', 20, 'MarkerFaceColor', [1 .6 .6], '
    MarkerEdgeColor', 'blue')
21 ylabel('Etiqueta de MAD', 'FontSize', 30, 'FontWeight', 'bold')
22 xlabel('Numero de entrada, N', 'FontSize', 30, 'FontWeight', 'bold
    ')
23 set(gca, 'FontSize', 24, 'linewidth', 3) ; axis tight
```

# Apéndice B

## Presentaciones en eventos académicos



Figura B.1: Constancia de participación en el tercer foro de avances estudiantiles, en la Red Temática CONACyT de Ciencias Forenses, C.U., 11/sept/2018.



## Propuesta metodológica para utilizar la ley de Benford en la investigación forense en resultados electorales

Óscar Jiménez Sánchez; [pqr7688@gmail.com](mailto:pqr7688@gmail.com) Facultad de Ciencias, UNAM  
 Vicente Torres Zúñiga; [vicentz@gmail.com](mailto:vicentz@gmail.com) Licenciatura en Ciencia Forense, UNAM

---

Resumen

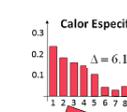
Exploramos la concordancia entre bases de datos electorales y ley de Benford, donde falla la prueba puede existir fraude. Presentamos un estándar matemático para interpretar y mostrar más claramente el dictamen. El comportamiento de la sucesión de Fibonacci es nuestro estándar imparcial, que permite homologar índices estadísticos. Contrastamos con datos empíricos; bases electorales presidenciales: México (2006 y 2018) y España (2015).

---

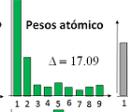
Introducción

Ley del primer dígito

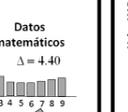
Es una ley de probabilidad que describe las frecuencias de la aparición del primer dígito (1,2,...,9). Muchos registros (naturales o no) presentan este comportamiento.



Calor Especifico  
 $\Delta = 6.10$



Pesos atómico  
 $\Delta = 17.09$



Datos matemáticos  
 $\Delta = 4.40$



Datos Combinados  
 $\Delta = 0.89$



Mark Negrini

$$P(d_i) = \sum_{k=1}^9 \log_{10} \left( 1 + \frac{1}{10k + d_i} \right)$$

Ref.: Walter R. Mebane, Jr., Election Forensics (2018).

Sucesión de Fibonacci

Es un arreglo de números donde cada término es el resultado de la suma de los dos anteriores. Permite modelar fenómenos naturales.

$f_0 = 0$   
 $y$   
 $f_1 = 1$



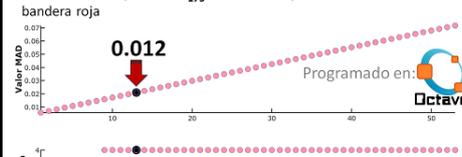
$$f_n = f_{n-1} + f_{n-2}$$

Ref.: M. Sambridge, H. Tkalcic, A. Jackson, Benford's law in the natural sciences (2010).

Metodología

Estándar propuesto

Sustitución de 1 por 9, en  $f_{175}$ . Se resalta el punto donde el valor es una bandera roja



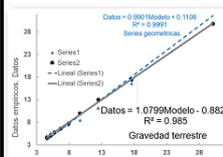
0.012

Programado en: 

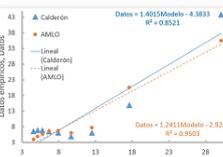
$$MAD = \sum_{i=1}^9 \frac{|\text{Observado}(i) - \text{Teoría}(i)|}{i}$$

Resultados

Comparaciones



Datos = 0.9901 Modelo = 0.1105  
 $R^2 = 0.9991$   
Series exponencial



Datos = 1.4015 Modelo = 4.3833  
 $R^2 = 0.8521$

Comparación por linealización entre bases de datos y la prueba teórica. El caso ideal es:  $y = 1x + 0$ ,  $R^2 = 1$ .

Elecciones presidenciales

México 2006:

España 2015:

México 2018:

Conclusiones

- Homologamos índices estadísticos de congruencia dato-teoría: 1) MAD, 2)  $\chi^2$ , 3) Prueba-Z 4) Correlación, 5) Linealización.
- Los códigos están AHORA a disposición de quien lo necesite, licencia: *creative commons*.
- Se necesita progresar en el algoritmo para analizar los dos primeros dígitos: 10, 11, 12, ..., 99.

Figura B.2: Cartel presentado en el tercer foro de avances estudiantiles, en la Red Temática CONACyT de Ciencias Forenses, C.U., 11/sept/2018.

## Propuesta sometida al «I Simposio Estudiantil de Estadística Forense», febrero 2020.



Figura B.3: Constancia de participación el primer simposio estudiantil de estadística forense.

### La sucesión de Fibonacci para equiparar criterios estadísticos ante la alerta de fraude

**Resumen** Las controversias sobre la integridad de elecciones pueden desencadenar violencia. Entonces, muchos auditores forenses emplean métodos estadísticos para alertar sobre alguna irregularidad electoral. Verbigracia, la popular ley de Benford; afirmando que la frecuencia en la aparición del primer dígito entre registros contables se comporta como una función logarítmica, donde el signo uno es el más frecuente y el nueve es más atípico. Empero, es inexistente la teorización de la conducta estadística de una elección limpia; así, los umbrales de alerta de fraude se sustentan en conjeturas subjetivas. Proponemos utilizar como modelo de comportamiento estadístico limpio a

---

una serie matemática exitosa en describir fenómenos naturales y sociales: la sucesión de Fibonacci; la que después alteramos con control y estrategia para evidenciar el significado de una tendencia fraudulenta. Esta perspectiva permite homologar criterios y comprensión en la interpretación de índices estadísticos; es decir, es de mejor comprensión para el público en general y en especial para los profesionales del derecho. Finalmente, con estos elementos analizamos bases de datos de elecciones presidenciales, algunas calificadas como limpias y otras como fraudulentas: México-2006, España-2015 y México-2018.

# Apéndice C

## Tablas logarítmicas de Newcomb

A manera de mostrar el deterioro en las páginas de tablas logarítmicas que inspiraron a Newcomb a formular la ley de Benford, presentamos aquí cinco páginas seleccionadas del libro: «*Logarithmic and other mathematical tables: with examples of their use and hints on the art of computation*». Se incluye la portada, las dos primeras páginas de las tablas de logaritmos comunes (que pese a la restauración muestran deterioro en la parte inferior izquierda) y las dos últimas de tal sección (carentes del deterioro mencionado).

- Autor: Newcomb, Simon, 1835-1909.
- Lenguaje: English
- Publicado por: New York : H. Holt, 1882.
- Descripción física : 80, 104 p. ; 23 cm.
- Versión impresa: Universidad de Harvard University
- Digitalizado por: Google
- Vinculo web: *Hathi trust*, Biblioteca digital.
- Último acceso: 20/enero/2020.

Schuyler B. Sevier  
July 14, 1899  
(General Reference)

---

NEWCOMB'S MATHEMATICAL COURSE

LOGARITHMIC AND OTHER  
MATHEMATICAL TABLES

WITH EXAMPLES OF THEIR USE AND HINTS ON THE ART OF  
COMPUTATION

BY

SIMON NEWCOMB

*Professor of Mathematics, in the Johns Hopkins University.*



NEW YORK  
HENRY HOLT AND COMPANY

TABLE I.  
—  
COMMON LOGARITHMS  
OF NUMBERS.

| N. | Log.       | N. | Log.     | N. | Log.     | N.  | Log.     | N.  | Log.     |
|----|------------|----|----------|----|----------|-----|----------|-----|----------|
| 0  | —Infinity. | 80 | 1.47 712 | 60 | 1.77 815 | 90  | 1.95 424 | 120 | 2.07 918 |
| 1  | 0.00 000   | 31 | 1.49 136 | 61 | 1.78 533 | 91  | 1.95 904 | 121 | 2.08 279 |
| 2  | 0.30 103   | 32 | 1.50 515 | 62 | 1.79 239 | 92  | 1.96 379 | 122 | 2.08 636 |
| 3  | 0.47 712   | 33 | 1.51 851 | 63 | 1.79 934 | 93  | 1.96 848 | 123 | 2.08 991 |
| 4  | 0.60 206   | 34 | 1.53 148 | 64 | 1.80 618 | 94  | 1.97 313 | 124 | 2.09 342 |
| 5  | 0.69 897   | 35 | 1.54 407 | 65 | 1.81 291 | 95  | 1.97 772 | 125 | 2.09 691 |
| 6  | 0.77 815   | 36 | 1.55 630 | 66 | 1.81 954 | 96  | 1.98 227 | 126 | 2.10 037 |
| 7  | 0.84 510   | 37 | 1.56 820 | 67 | 1.82 607 | 97  | 1.98 677 | 127 | 2.10 380 |
| 8  | 0.90 309   | 38 | 1.57 978 | 68 | 1.83 251 | 98  | 1.99 123 | 128 | 2.10 721 |
| 9  | 0.95 424   | 39 | 1.59 106 | 69 | 1.83 885 | 99  | 1.99 564 | 129 | 2.11 059 |
| 10 | 1.00 000   | 40 | 1.60 206 | 70 | 1.84 510 | 100 | 2.00 000 | 130 | 2.11 394 |
| 11 | 1.04 139   | 41 | 1.61 278 | 71 | 1.85 126 | 101 | 2.00 432 | 131 | 2.11 727 |
| 12 | 1.07 918   | 42 | 1.62 325 | 72 | 1.85 733 | 102 | 2.00 860 | 132 | 2.12 057 |
| 13 | 1.11 394   | 43 | 1.63 347 | 73 | 1.86 332 | 103 | 2.01 284 | 133 | 2.12 385 |
| 14 | 1.14 613   | 44 | 1.64 345 | 74 | 1.86 923 | 104 | 2.01 703 | 134 | 2.12 710 |
| 15 | 1.17 609   | 45 | 1.65 321 | 75 | 1.87 506 | 105 | 2.02 119 | 135 | 2.13 033 |
| 16 | 1.20 412   | 46 | 1.66 276 | 76 | 1.88 081 | 106 | 2.02 531 | 136 | 2.13 354 |
| 17 | 1.23 045   | 47 | 1.67 210 | 77 | 1.88 649 | 107 | 2.02 938 | 137 | 2.13 672 |
| 18 | 1.25 527   | 48 | 1.68 124 | 78 | 1.89 209 | 108 | 2.03 342 | 138 | 2.13 988 |
| 19 | 1.27 875   | 49 | 1.69 020 | 79 | 1.89 763 | 109 | 2.03 743 | 139 | 2.14 301 |
| 20 | 1.30 103   | 50 | 1.69 897 | 80 | 1.90 309 | 110 | 2.04 139 | 140 | 2.14 613 |
| 21 | 1.32 222   | 51 | 1.70 757 | 81 | 1.90 849 | 111 | 2.04 532 | 141 | 2.14 922 |
| 22 | 1.34 242   | 52 | 1.71 600 | 82 | 1.91 381 | 112 | 2.04 922 | 142 | 2.15 229 |
| 23 | 1.36 173   | 53 | 1.72 428 | 83 | 1.91 908 | 113 | 2.05 308 | 143 | 2.15 534 |
| 24 | 1.38 021   | 54 | 1.73 239 | 84 | 1.92 428 | 114 | 2.05 690 | 144 | 2.15 836 |
| 25 | 1.39 794   | 55 | 1.74 036 | 85 | 1.92 942 | 115 | 2.06 070 | 145 | 2.16 137 |
| 26 | 1.41 497   | 56 | 1.74 819 | 86 | 1.93 450 | 116 | 2.06 446 | 146 | 2.16 435 |
| 27 | 1.43 136   | 57 | 1.75 587 | 87 | 1.93 952 | 117 | 2.06 819 | 147 | 2.16 732 |
| 28 | 1.44 716   | 58 | 1.76 343 | 88 | 1.94 448 | 118 | 2.07 188 | 148 | 2.17 026 |
| 29 | 1.46 240   | 59 | 1.77 085 | 89 | 1.94 939 | 119 | 2.07 555 | 149 | 2.17 319 |
| 80 | 1.47 712   | 60 | 1.77 815 | 90 | 1.95 424 | 120 | 2.07 918 | 150 | 2.17 609 |

Generated on 2020-01-22 14:44 GMT / http://hdl.handle.net/2027/hvd.hn5zmb  
Public Domain, Google-digitized / http://www.hathitrust.org/access\_use#pd-google

TABLE I.

| N.  | 0      | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | Prop. Pts.       |
|-----|--------|------|------|------|------|------|------|------|------|------|------------------|
| 100 | 00 000 | 043  | 087  | 130  | 173  | 217  | 260  | 303  | 346  | 389  |                  |
| 01  | 432    | 475  | 518  | 561  | 604  | 647  | 689  | 732  | 775  | 817  |                  |
| 02  | 860    | 903  | 945  | 988  | *030 | *072 | *115 | *157 | *199 | *242 | 44 43 42         |
| 03  | 01 284 | 326  | 368  | 410  | 452  | 494  | 536  | 578  | 620  | 662  | 1 4.4 4 3 4.2    |
| 04  | 703    | 745  | 787  | 828  | 870  | 912  | 953  | 995  | *036 | *078 | 2 8 8 8 6 8.4    |
| 05  | 02 119 | 160  | 202  | 243  | 284  | 325  | 366  | 407  | 449  | 490  | 3 13.2 12 9 12 6 |
| 06  | 531    | 572  | 612  | 653  | 694  | 735  | 776  | 816  | 857  | 898  | 4 17 6 17 2 16 8 |
| 07  | 938    | 979  | *019 | *060 | *100 | *141 | *181 | *222 | *262 | *302 | 5 22 0 21 5 21.0 |
| 08  | 03 342 | 383  | 423  | 463  | 503  | 543  | 583  | 623  | 663  | 703  | 6 26 4 25 8 25.2 |
| 09  | 743    | 782  | 822  | 862  | 902  | 941  | 981  | *021 | *060 | *100 | 7 30 8 30.1 29 4 |
| 110 | 04 139 | 179  | 218  | 258  | 297  | 336  | 376  | 415  | 454  | 493  | 8 35 2 34 4 33 6 |
| 11  | 532    | 571  | 610  | 650  | 689  | 727  | 766  | 805  | 844  | 883  | 9 39 6 38 7 37 8 |
| 12  | 922    | 961  | 999  | *038 | *077 | *115 | *154 | *192 | *231 | *269 | 41 40 39         |
| 13  | 05 308 | 346  | 385  | 423  | 461  | 500  | 538  | 576  | 614  | 652  | 1 4.1 4.0 3 9    |
| 14  | 690    | 729  | 767  | 805  | 843  | 881  | 918  | 956  | 994  | *032 | 2 8 2 8 0 7 8    |
| 15  | 06 070 | 108  | 145  | 183  | 221  | 258  | 296  | 333  | 371  | 408  | 3 12 3 12 0 11 7 |
| 16  | 446    | 483  | 521  | 558  | 595  | 633  | 670  | 707  | 744  | 781  | 4 16 4 16 0 15 6 |
| 17  | 819    | 856  | 893  | 930  | 967  | *004 | *041 | *078 | *115 | *151 | 5 20 5 20 0 19 5 |
| 18  | 07 188 | 225  | 262  | 298  | 335  | 372  | 408  | 445  | 482  | 518  | 6 24 6 24 0 23 4 |
| 19  | 555    | 591  | 628  | 664  | 700  | 737  | 773  | 809  | 846  | 882  | 7 28 7 28 0 27 3 |
| 120 | 918    | 954  | 990  | *027 | *063 | *099 | *135 | *171 | *207 | *243 | 8 32 8 32 0 31 2 |
| 21  | 08 270 | 314  | 350  | 386  | 422  | 458  | 493  | 529  | 565  | 600  | 9 36 9 36 0 35 1 |
| 22  | 636    | 672  | 707  | 743  | 778  | 814  | 849  | 884  | 920  | 955  | 38 37 36         |
| 23  | 991    | *026 | *061 | *096 | *132 | *167 | *202 | *237 | *272 | *307 | 1 3 8 3 7 3 6    |
| 24  | 09 342 | 377  | 412  | 447  | 482  | 517  | 552  | 587  | 621  | 656  | 2 7 6 7 4 7 2    |
| 25  | 691    | 726  | 760  | 795  | 830  | 864  | 899  | 934  | 968  | *003 | 3 11 4 11 3 10 8 |
| 26  | 10 037 | 072  | 106  | 140  | 175  | 209  | 243  | 278  | 312  | 346  | 4 15 2 14 8 14 4 |
| 27  | 380    | 415  | 449  | 483  | 517  | 551  | 585  | 619  | 653  | 687  | 5 19 0 18 5 18 0 |
| 28  | 721    | 755  | 789  | 823  | 857  | 890  | 924  | 958  | 992  | *025 | 6 22 8 22 2 21 6 |
| 29  | 11 059 | 093  | 126  | 160  | 193  | 227  | 261  | 294  | 327  | 361  | 7 26 6 25 9 25 2 |
| 180 | 394    | 428  | 461  | 494  | 528  | 561  | 594  | 628  | 661  | 694  | 8 30 4 29 6 28 8 |
| 31  | 727    | 760  | 793  | 826  | 860  | 893  | 926  | 959  | 992  | *024 | 9 34 2 33 3 32 4 |
| 32  | 12 057 | 090  | 123  | 156  | 189  | 222  | 254  | 287  | 320  | 352  | 35 34 33         |
| 33  | 385    | 418  | 450  | 483  | 516  | 548  | 581  | 613  | 646  | 678  | 1 3 5 3 4 3 3    |
| 34  | 710    | 743  | 775  | 808  | 840  | 872  | 905  | 937  | 969  | *001 | 2 7 0 6 8 6 6    |
| 35  | 13 033 | 066  | 098  | 130  | 162  | 194  | 226  | 258  | 290  | 322  | 3 10 5 10 3 9 9  |
| 36  | 354    | 386  | 418  | 450  | 481  | 513  | 545  | 577  | 609  | 640  | 4 14 0 13 6 13 2 |
| 37  | 672    | 704  | 735  | 767  | 799  | 830  | 862  | 893  | 925  | 956  | 5 17 5 17 0 16 5 |
| 38  | 988    | *019 | *051 | *082 | *114 | *145 | *176 | *208 | *239 | *270 | 6 21 0 20 4 19 8 |
| 39  | 14 301 | 333  | 364  | 395  | 426  | 457  | 489  | 520  | 551  | 582  | 7 24 5 23 6 23 1 |
| 140 | 613    | 644  | 675  | 706  | 737  | 768  | 799  | 829  | 860  | 891  | 8 28 0 27 2 26 4 |
| 41  | 922    | 953  | 983  | *014 | *045 | *076 | *106 | *137 | *168 | *198 | 9 31 5 30 6 29 7 |
| 42  | 15 229 | 259  | 290  | 320  | 351  | 381  | 412  | 442  | 473  | 503  | 38 37 36         |
| 43  | 534    | 564  | 594  | 625  | 655  | 685  | 715  | 746  | 776  | 806  | 1 3 2 3 1 3 0    |
| 44  | 836    | 866  | 897  | 927  | 957  | 987  | *017 | *047 | *077 | *107 | 2 6 4 6 2 6 0    |
| 45  | 16 137 | 167  | 197  | 227  | 256  | 286  | 316  | 346  | 376  | 406  | 3 9 6 9 3 9 0    |
| 46  | 435    | 465  | 495  | 524  | 554  | 584  | 613  | 643  | 673  | 702  | 4 12 8 12 4 12 0 |
| 47  | 732    | 761  | 791  | 820  | 850  | 879  | 909  | 938  | 967  | 997  | 5 16 0 15 5 15 0 |
| 48  | 17 026 | 056  | 085  | 114  | 143  | 173  | 202  | 231  | 260  | 289  | 6 19 2 18 6 18 0 |
| 49  | 319    | 348  | 377  | 406  | 435  | 464  | 493  | 522  | 551  | 580  | 7 22 4 21 7 21 0 |
| 150 | 609    | 638  | 667  | 696  | 725  | 754  | 782  | 811  | 840  | 869  | 8 25 6 24 8 24 0 |
| 918 | 954    | 990  | *027 | *063 | *099 | *135 | *171 | *207 | *243 |      | 9 28 8 27 9 27 0 |

Generated on 2020-01-22 14:44 GMT / http://hdl.handle.net/2027/hvd.hn5zmb  
Public Domain, Google-digitized / http://www.hathitrust.org/access\_use#pd-google

N. 0

Go gle

Harvard University

20

TABLE I.

| N.          | 0   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | Prop. Pts. |     |      |
|-------------|-----|------|------|------|------|------|------|------|------|------|------------|-----|------|
| <b>1000</b> | 000 | 000  | 043  | 087  | 130  | 174  | 217  | 260  | 304  | 347  | 391        |     |      |
| 1001        | 434 | 477  | 521  | 564  | 608  | 651  | 694  | 738  | 781  | 824  |            |     |      |
| 1002        | 868 | 911  | 954  | 998  | *041 | *084 | *128 | *171 | *214 | *258 |            |     |      |
| 1003        | 001 | 301  | 344  | 388  | 431  | 474  | 517  | 561  | 604  | 647  | 690        | 44  |      |
| 1004        | 734 | 777  | 820  | 863  | 907  | 950  | 993  | *036 | *080 | *123 | 1          | 4.4 |      |
| 1005        | 002 | 166  | 209  | 252  | 296  | 339  | 382  | 425  | 468  | 512  | 555        | 2   | 8.8  |
| 1006        | 598 | 641  | 684  | 727  | 771  | 814  | 857  | 900  | 943  | 986  |            | 3   | 13.2 |
| 1007        | 003 | 029  | 073  | 116  | 159  | 202  | 245  | 288  | 331  | 374  | 417        | 4   | 17.6 |
| 1008        | 461 | 504  | 547  | 590  | 633  | 676  | 719  | 762  | 805  | 848  |            | 5   | 22.0 |
| 1009        | 891 | 934  | 977  | *020 | *063 | *106 | *149 | *192 | *235 | *278 |            | 6   | 26.4 |
| <b>1010</b> | 004 | 321  | 364  | 407  | 450  | 493  | 536  | 579  | 622  | 665  | 708        | 7   | 30.8 |
| 1011        | 751 | 794  | 837  | 880  | 923  | 966  | *009 | *052 | *095 | *138 |            | 8   | 35.2 |
| 1012        | 005 | 180  | 223  | 266  | 309  | 352  | 395  | 438  | 481  | 524  | 567        | 9   | 39.6 |
| 1013        | 609 | 652  | 695  | 738  | 781  | 824  | 867  | 909  | 952  | 995  |            |     |      |
| 1014        | 006 | 038  | 081  | 124  | 166  | 209  | 252  | 295  | 338  | 380  | 423        |     |      |
| 1015        | 466 | 509  | 552  | 594  | 637  | 680  | 723  | 765  | 808  | 851  |            |     |      |
| 1016        | 894 | 936  | 979  | *022 | *065 | *107 | *150 | *193 | *236 | *278 |            | 1   | 4.3  |
| 1017        | 007 | 321  | 364  | 406  | 449  | 492  | 534  | 577  | 620  | 662  | 705        | 2   | 8.6  |
| 1018        | 748 | 790  | 833  | 876  | 918  | 961  | *004 | *046 | *089 | *132 |            | 3   | 12.9 |
| 1019        | 008 | 174  | 217  | 259  | 302  | 345  | 387  | 430  | 472  | 515  | 558        | 4   | 17.2 |
| <b>1020</b> | 600 | 643  | 685  | 728  | 770  | 813  | 856  | 898  | 941  | 983  |            | 5   | 21.5 |
| 1021        | 009 | 026  | 068  | 111  | 153  | 196  | 238  | 281  | 323  | 366  | 408        | 6   | 25.8 |
| 1022        | 451 | 493  | 536  | 578  | 621  | 663  | 706  | 748  | 791  | 833  |            | 7   | 30.1 |
| 1023        | 876 | 918  | 961  | *003 | *045 | *088 | *130 | *173 | *215 | *258 |            | 8   | 34.4 |
| 1024        | 010 | 300  | 342  | 385  | 427  | 470  | 512  | 554  | 597  | 639  | 681        | 9   | 38.7 |
| 1025        | 724 | 766  | 809  | 851  | 893  | 936  | 978  | *020 | *063 | *105 |            |     |      |
| 1026        | 011 | 147  | 190  | 232  | 274  | 317  | 359  | 401  | 444  | 486  | 528        |     |      |
| 1027        | 570 | 613  | 655  | 697  | 740  | 782  | 824  | 866  | 909  | 951  |            |     |      |
| 1028        | 993 | *035 | *078 | *120 | *162 | *204 | *247 | *289 | *331 | *373 |            | 1   | 4.2  |
| 1029        | 012 | 415  | 458  | 500  | 542  | 584  | 626  | 669  | 711  | 753  | 795        | 2   | 8.4  |
| <b>1030</b> | 837 | 879  | 922  | 964  | *006 | *048 | *090 | *132 | *174 | *217 |            | 3   | 12.6 |
| 1031        | 013 | 259  | 301  | 343  | 385  | 427  | 469  | 511  | 553  | 596  | 638        | 4   | 16.8 |
| 1032        | 680 | 722  | 764  | 806  | 848  | 890  | 932  | 974  | *016 | *058 |            | 5   | 21.0 |
| 1033        | 014 | 100  | 142  | 184  | 226  | 268  | 310  | 352  | 395  | 437  | 479        | 6   | 25.2 |
| 1034        | 521 | 563  | 605  | 647  | 689  | 730  | 772  | 814  | 856  | 898  |            | 7   | 29.4 |
| 1035        | 940 | 982  | *024 | *066 | *108 | *150 | *192 | *234 | *276 | *318 |            | 8   | 33.6 |
| 1036        | 015 | 360  | 402  | 444  | 485  | 527  | 569  | 611  | 653  | 695  | 737        | 9   | 37.8 |
| 1037        | 779 | 821  | 863  | 904  | 946  | 988  | *030 | *072 | *114 | *156 |            |     |      |
| 1038        | 016 | 197  | 239  | 281  | 323  | 365  | 407  | 448  | 490  | 532  | 574        |     |      |
| 1039        | 616 | 657  | 699  | 741  | 783  | 824  | 866  | 908  | 950  | 992  |            |     |      |
| <b>1040</b> | 017 | 033  | 075  | 117  | 159  | 200  | 242  | 284  | 326  | 367  | 409        | 1   | 4.1  |
| 1041        | 451 | 492  | 534  | 576  | 618  | 659  | 701  | 743  | 784  | 826  |            | 2   | 8.2  |
| 1042        | 868 | 909  | 951  | 993  | *034 | *076 | *118 | *159 | *201 | *243 |            | 3   | 12.3 |
| 1043        | 018 | 284  | 326  | 368  | 409  | 451  | 492  | 534  | 576  | 617  | 659        | 4   | 16.4 |
| 1044        | 700 | 742  | 784  | 825  | 867  | 908  | 950  | 992  | *033 | *075 |            | 5   | 20.5 |
| 1045        | 019 | 116  | 158  | 199  | 241  | 282  | 324  | 366  | 407  | 449  | 490        | 6   | 24.6 |
| 1046        | 532 | 573  | 615  | 656  | 698  | 739  | 781  | 822  | 864  | 905  |            | 7   | 28.7 |
| 1047        | 947 | 988  | *030 | *071 | *113 | *154 | *195 | *237 | *278 | *320 |            | 8   | 32.8 |
| 1048        | 020 | 361  | 403  | 444  | 486  | 527  | 568  | 610  | 651  | 693  | 734        | 9   | 36.9 |
| 1049        | 775 | 817  | 858  | 900  | 941  | 982  | *024 | *065 | *107 | *148 |            |     |      |
| <b>1050</b> | 021 | 189  | 231  | 272  | 313  | 355  | 396  | 437  | 479  | 520  | 561        |     |      |
| N.          | 0   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | Prop. Pts. |     |      |

Generated on 2020-01-22 14:45 GMT / http://hdl.handle.net/2027/hvd.hn5zmb  
Public Domain, Google-digitized / http://www.hathitrust.org/access\_use#pd-google

LOGARITHMS OF NUMBERS.

21

| N.          | 0       | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | Prop. Pts. |
|-------------|---------|------|------|------|------|------|------|------|------|------|------------|
| <b>1050</b> | 021 189 | 231  | 272  | 313  | 355  | 396  | 437  | 479  | 520  | 561  |            |
| 1051        | 603     | 644  | 685  | 727  | 768  | 809  | 851  | 892  | 933  | 974  |            |
| 1052        | 022 016 | 057  | 098  | 140  | 181  | 222  | 263  | 305  | 346  | 387  |            |
| 1053        | 428     | 470  | 511  | 552  | 593  | 635  | 676  | 717  | 758  | 799  | 42         |
| 1054        | 841     | 882  | 923  | 964  | *005 | *047 | *088 | *129 | *170 | *211 | 1 4.2      |
| 1055        | 023 252 | 294  | 335  | 376  | 417  | 458  | 499  | 541  | 582  | 623  | 2 8.4      |
| 1056        | 664     | 705  | 746  | 787  | 828  | 870  | 911  | 952  | 993  | *034 | 3 12.6     |
| 1057        | 024 075 | 116  | 157  | 198  | 239  | 280  | 321  | 363  | 404  | 445  | 4 16.8     |
| 1058        | 486     | 527  | 568  | 609  | 650  | 691  | 732  | 773  | 814  | 855  | 5 21.0     |
| 1059        | 896     | 937  | 978  | *019 | *060 | *101 | *142 | *183 | *224 | *265 | 6 25.2     |
| <b>1060</b> | 025 306 | 347  | 388  | 429  | 470  | 511  | 552  | 593  | 634  | 674  | 7 29.4     |
| 1061        | 715     | 756  | 797  | 838  | 879  | 920  | 961  | *002 | *043 | *084 | 8 33.6     |
| 1062        | 026 125 | 165  | 206  | 247  | 288  | 329  | 370  | 411  | 452  | 492  | 9 37.8     |
| 1063        | 533     | 574  | 615  | 656  | 697  | 737  | 778  | 819  | 860  | 901  |            |
| 1064        | 942     | 982  | *023 | *064 | *105 | *146 | *186 | *227 | *268 | *309 |            |
| 1065        | 027 350 | 390  | 431  | 472  | 513  | 553  | 594  | 635  | 676  | 716  | 41         |
| 1066        | 757     | 798  | 839  | 879  | 920  | 961  | *002 | *042 | *083 | *124 | 1 4.1      |
| 1067        | 028 164 | 205  | 246  | 287  | 327  | 368  | 409  | 449  | 490  | 531  | 2 8.2      |
| 1068        | 571     | 612  | 653  | 693  | 734  | 775  | 815  | 856  | 896  | 937  | 3 12.3     |
| 1069        | 978     | *018 | *059 | *100 | *140 | *181 | *221 | *262 | *303 | *343 | 4 16.4     |
| <b>1070</b> | 029 384 | 424  | 465  | 506  | 546  | 587  | 627  | 668  | 708  | 749  | 5 20.5     |
| 1071        | 789     | 830  | 871  | 911  | 952  | 992  | *033 | *073 | *114 | *154 | 6 24.6     |
| 1072        | 030 195 | 235  | 276  | 316  | 357  | 397  | 438  | 478  | 519  | 559  | 7 28.7     |
| 1073        | 600     | 640  | 681  | 721  | 762  | 802  | 843  | 883  | 923  | 964  | 8 32.8     |
| 1074        | 031 004 | 045  | 085  | 126  | 166  | 206  | 247  | 287  | 328  | 368  | 9 36.9     |
| 1075        | 408     | 449  | 489  | 530  | 570  | 610  | 651  | 691  | 732  | 772  |            |
| 1076        | 812     | 853  | 893  | 933  | 974  | *014 | *054 | *095 | *135 | *175 |            |
| 1077        | 032 216 | 256  | 296  | 337  | 377  | 417  | 458  | 498  | 538  | 578  | 1 4.0      |
| 1078        | 619     | 659  | 699  | 740  | 780  | 820  | 860  | 901  | 941  | 981  | 2 8.0      |
| 1079        | 033 021 | 062  | 102  | 142  | 182  | 223  | 263  | 303  | 343  | 384  | 3 12.0     |
| <b>1080</b> | 424     | 464  | 504  | 544  | 585  | 625  | 665  | 705  | 745  | 785  | 4 16.0     |
| 1081        | 826     | 866  | 906  | 946  | 986  | *027 | *067 | *107 | *147 | *187 | 5 20.0     |
| 1082        | 034 227 | 267  | 308  | 348  | 388  | 428  | 468  | 508  | 548  | 588  | 6 24.0     |
| 1083        | 628     | 669  | 709  | 749  | 789  | 829  | 869  | 909  | 949  | 989  | 7 28.0     |
| 1084        | 035 029 | 069  | 109  | 149  | 190  | 230  | 270  | 310  | 350  | 390  | 8 32.0     |
| 1085        | 430     | 470  | 510  | 550  | 590  | 630  | 670  | 710  | 750  | 790  | 9 36.0     |
| 1086        | 830     | 870  | 910  | 950  | 990  | *030 | *070 | *110 | *150 | *190 |            |
| 1087        | 036 230 | 269  | 309  | 349  | 389  | 429  | 469  | 509  | 549  | 589  |            |
| 1088        | 629     | 669  | 709  | 749  | 789  | 828  | 868  | 908  | 948  | 988  |            |
| 1089        | 037 028 | 068  | 108  | 148  | 187  | 227  | 267  | 307  | 347  | 387  | 89         |
| <b>1090</b> | 426     | 466  | 506  | 546  | 586  | 626  | 665  | 705  | 745  | 785  | 1 3.9      |
| 1091        | 825     | 865  | 904  | 944  | 984  | *024 | *064 | *103 | *143 | *183 | 2 7.8      |
| 1092        | 038 223 | 262  | 302  | 342  | 382  | 421  | 461  | 501  | 541  | 580  | 3 11.7     |
| 1093        | 620     | 660  | 700  | 739  | 779  | 819  | 859  | 898  | 938  | 978  | 4 15.6     |
| 1094        | 039 017 | 057  | 097  | 136  | 176  | 216  | 255  | 295  | 335  | 374  | 5 19.5     |
| 1095        | 414     | 454  | 493  | 533  | 573  | 612  | 652  | 692  | 731  | 771  | 6 23.4     |
| 1096        | 811     | 850  | 890  | 929  | 969  | *009 | *048 | *088 | *127 | *167 | 7 27.3     |
| 1097        | 040 207 | 246  | 286  | 325  | 365  | 405  | 444  | 484  | 523  | 563  | 8 31.2     |
| 1098        | 602     | 642  | 681  | 721  | 761  | 800  | 840  | 879  | 919  | 958  | 9 35.1     |
| 1099        | 998     | *037 | *077 | *116 | *156 | *195 | *235 | *274 | *314 | *353 |            |
| <b>1100</b> | 041 393 | 432  | 472  | 511  | 551  | 590  | 630  | 669  | 708  | 748  |            |
| N.          | 0       | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | Prop. Pts. |

Generated on 2020-01-22 14:45 GMT / http://hdl.handle.net/2027/hvd.hn5zmb  
Public Domain, Google-digitized / http://www.hathitrust.org/access\_use#pd-google