

Evidencia de Manipulación Artificial de los Resultados al Aplicar la Ley de Benford al Referéndum Venezolano de Agosto 2004!

" !#\$!%& ' ()% " *+ , - -
. /*0)(- *121%#* ' 3 /%4 , 5602(7%8255)%1)%#2(9) /):2-%
;)<2(92 ') /9 , %1)%=6- *\$27%>2(2\$2-7%8) /)?@)52!

Qué es la Ley de Benford?

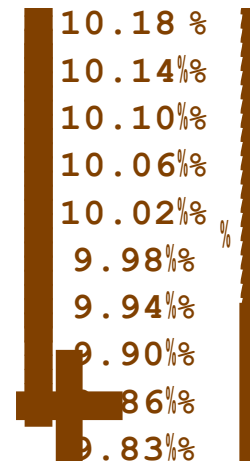
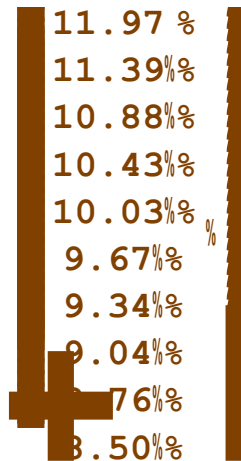
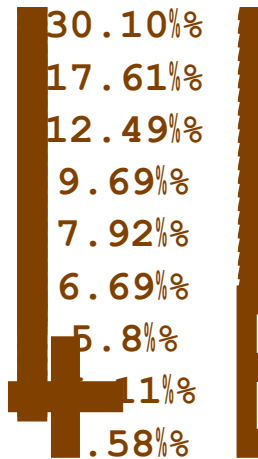
- Descubierta por Simon Newcomb (Astrónomo) en 1881 (tablas de logaritmos sucias).
- Redescubierta por el físico Frank Benford y publicada en el artículo: “La Ley de los Números Anómalos” (1938).
- “Ley” Empírica (en principio) y a la cual el Dr. Ted Hill dio cierta base teórica en 1996.
- Definición: Establece que con una “gran” frecuencia los números generados por procesos naturales medibles, no tienen una distribución uniforme en los dígitos de sus representaciones decimales. La distribución establecida empíricamente por Newcomb es logarítmica y depende de la posición decimal del dígito. En particular los números generados por los seres humanos presentan usualmente una distribución de Benford.
- Mark Nigrini la aplicó a la detección de fraude en las declaraciones de impuestos en 1993 (tesis de grado).
- Se publica un artículo divulgativo al respecto en la revista “Der Spiegel” en 1998.



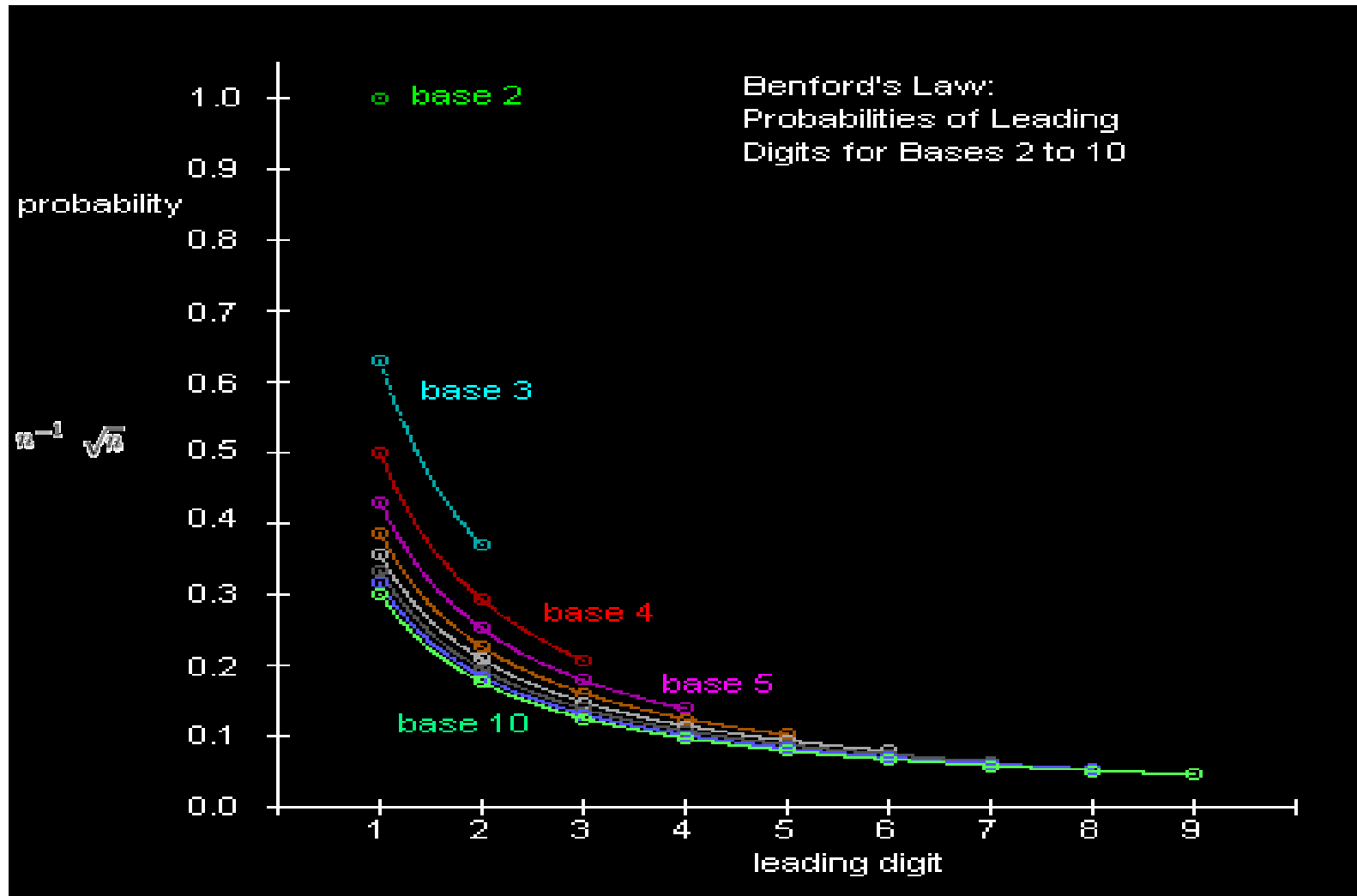
$$\text{Log } 1C \frac{1}{N}$$

$$iB1 \text{ Log } 1C \frac{1}{10iC N}$$

$$iB1 \quad jB0 \text{ Log } 1C \frac{1}{100iC 10jC N}$$



La Distribución de Benford es Invariante de Base y Escala.





Title	First Digit									Count
	1	2	3	4	5	6	7	8	9	
Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
Molec. Weight	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
Atomic Weight	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
n-1 , \sqrt{n} ,...	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
<i>Digest</i>	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
n1,n2 ,...n!	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Predicted	30.1	17.6	12.5	9.69	7.92	6.70	5.80	5.11	4.58	

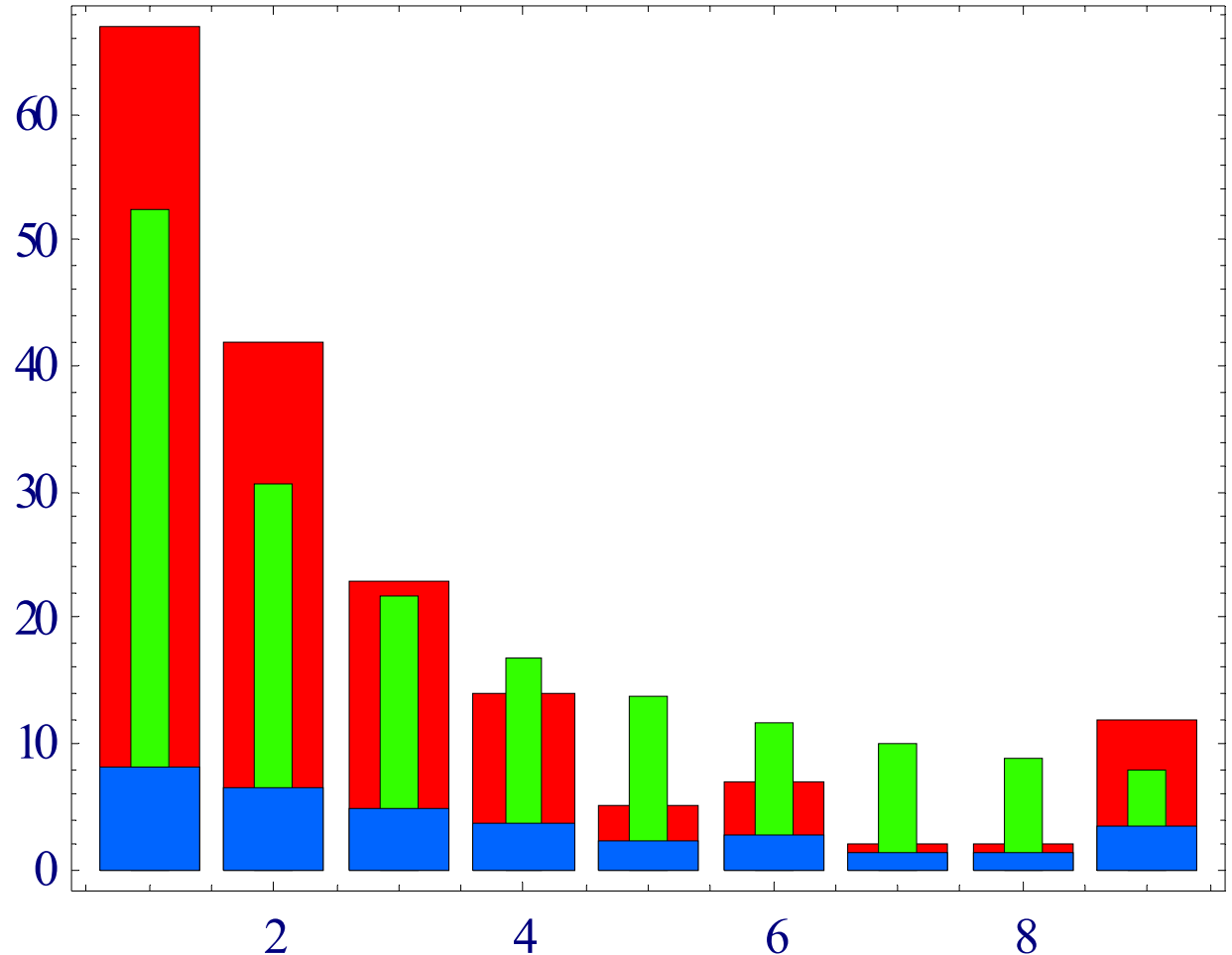
Números Recolectados por Benford.

Benford en las Olimpiadas.

Distribución del 1er Dígito del # de Medallas Olímpicas vs. Distribución de Benford

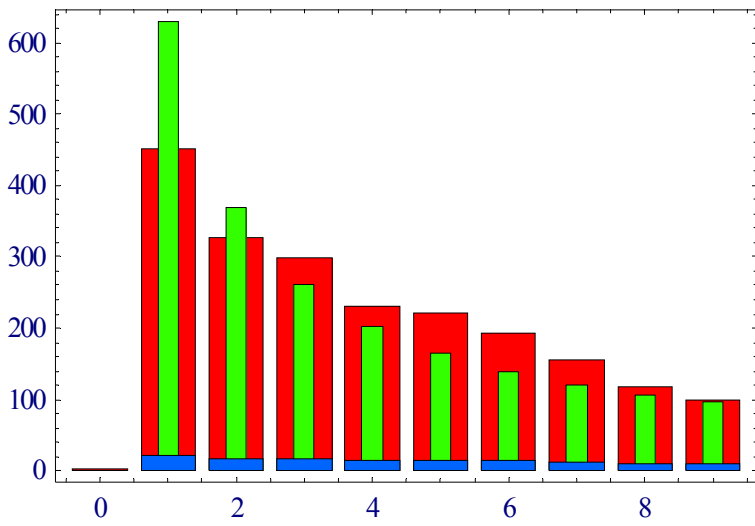
Rank by Gold	NOC	NOC Name				Tot.	Rank by Total
1	USA	United States	35	39	29	103	1
2	CHN	China	32	17	14	63	3
3	RUS	Russia	27	27	38	92	2
4	AUS	Australia	17	16	16	49	4
5	JPN	Japan	16	9	12	37	6
6	GER	Germany	14	16	18	48	5
7	FRA	France	11	9	13	33	7
8	ITA	Italy	10	11	11	32	8
9	KOR	Korea	9	12	9	30	=9
10	GBR	Great Britain	9	9	12	30	=9
11	CUB	Cuba	9	7	11	27	11
12	UKR	Ukraine	9	5	9	23	12
13	HUN	Hungary	8	6	3	17	16
14	ROM	Romania	8	5	6	19	=14
15	GRE	Greece	6	6	4	16	17
16	NOR	Norway	5	0	1	6	=32
17	NED	Netherlands	4	9	9	22	13
18	BRA	Brazil	4	3	3	10	=21
19	SWE	Sweden	4	1	2	7	=28
20	ESP	Spain	3	11	5	19	=14
21	CAN	Canada	3	6	3	12	=19
22	TUR	Turkey	3	3	4	10	=21
23	POL	Poland	3	2	5	10	=21
24	NZL	New Zealand	3	2	0	5	=37
25	THA	Thailand	3	1	4	8	=24
26	BLR	Belarus	2	6	7	15	18
27	AUT	Austria	2	4	1	7	=28
28	ETH	Ethiopia	2	3	2	7	=28
=29	IRI	I. R. Iran	2	2	2	6	=32
=29	SVK	Slovakia	2	2	2	6	=32
31	TPE	Chinese Taipei	2	2	1	5	=37
32	GEO	Georgia	2	2	0	4	=46

33	BUL	Bulgaria	2	1	9	12	=19
=34	JAM	Jamaica	2	1	2	5	=37
=34	UZB	Uzbekistan	2	1	2	5	=37
36	MAR	Morocco	2	1	0	3	=51
37	DEN	Denmark	2	0	6	8	=24
38	ARG	Argentina	2	0	4	6	=32
39	CHI	Chile	2	0	1	3	=51
40	KAZ	Kazakhstan	1	4	3	8	=24
41	KEN	Kenya	1	4	2	7	=28
42	CZE	Czech Republic	1	3	4	8	=24
43	RSA	South Africa	1	3	2	6	=32
44	CRO	Croatia	1	2	2	5	=37
45	LTU	Lithuania	1	2	0	3	=51
=46	EGY	Egypt	1	1	3	5	=37
=46	SUI	Switzerland	1	1	3	5	=37
48	INA	Indonesia	1	1	2	4	=46
49	ZIM	Zimbabwe	1	1	1	3	=51
50	AZE	Azerbaijan	1	0	4	5	=37
51	BEL	Belgium	1	0	2	3	=51
=52	BAH	Bahamas	1	0	1	2	=58
=52	ISR	Israel	1	0	1	2	=58
=54	CHR	Cameroon	1	0	0	1	=64
=54	DOM	Dominican Rep	1	0	0	1	=64
=54	IRL	Ireland	1	0	0	1	=64
=54	UAE	U Arab Emirates	1	0	0	1	=64
58	PRK	DPR Korea	0	4	1	5	=37
59	LAT	Latvia	0	4	0	4	=46
60	MEX	Mexico	0	3	1	4	=46
61	POR	Portugal	0	2	1	3	=51
=62	FIN	Finland	0	2	0	2	=58
=62	SCG	Serbia/ Monteneg	0	2	0	2	=58
64	SLO	Slovenia	0	1	3	4	=46
65	EST	Estonia	0	1	2	3	=51
=66	HKG	Hong Kong	0	1	0	1	=64
=66	IND	India	0	1	0	1	=64
=66	PAR	Paraguay	0	1	0	1	=64



Benford Útil

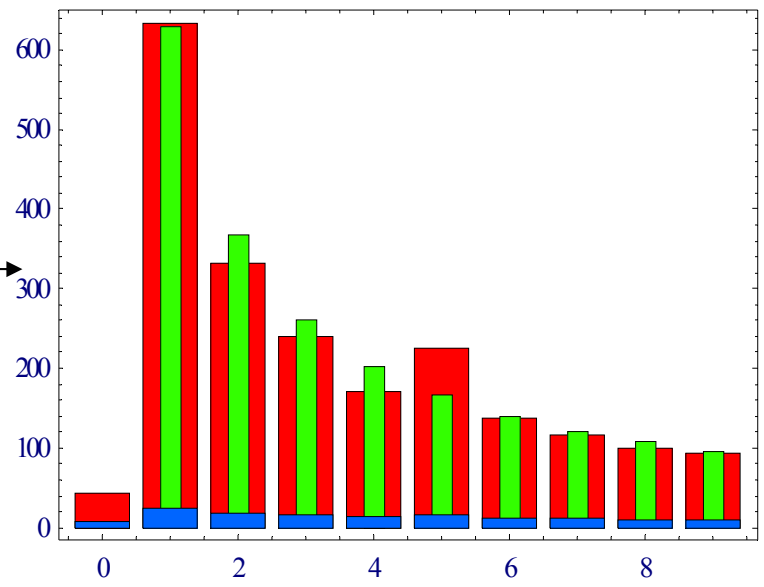
1er Dígito de la Resta del Tiempo de Corredores Sucesivos vs. Benford



Primer Gráfico: Tiempos del Chip.

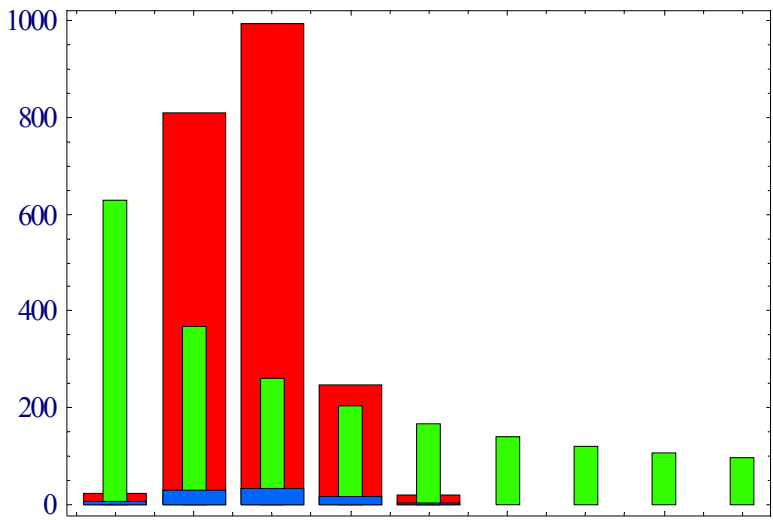
$$S = 9.8G 10^{H3}$$

1er Dígito de la Resta del Tiempo de Corredores Sucesivos vs. Benford



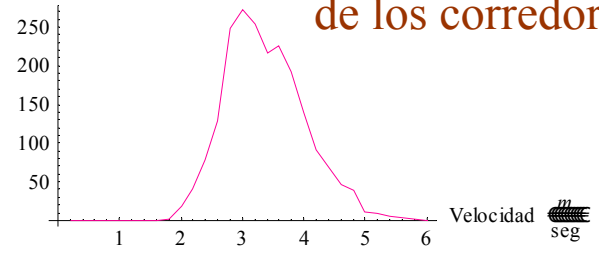
Segundo Gráfico: Tiempos Oficiales ó Tiempos del Chip Ordenados. $S = 1.5G 10^{H3}$

Distribución del Primer Dígito de las Velocidades vs. Benford



Tercer Gráfico: Velocidades promedio de los corredores.

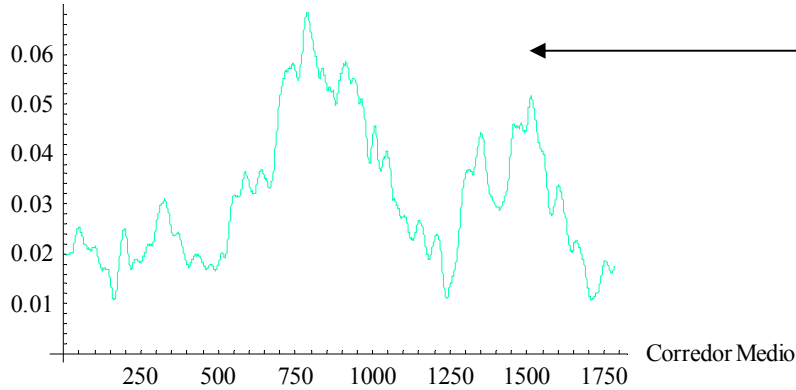
Número de Corredores



El Poder de Uno!

Midiendo la Voluntad Humana.

Ajuste de Distribución

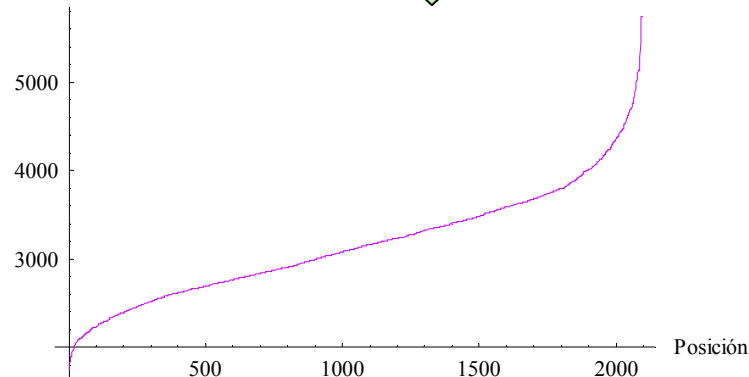
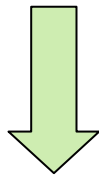


Primer Gráfico: Ajuste a la Distribución de Benford en función de la posición media de los corredores (subconjuntos de 300 corredores).

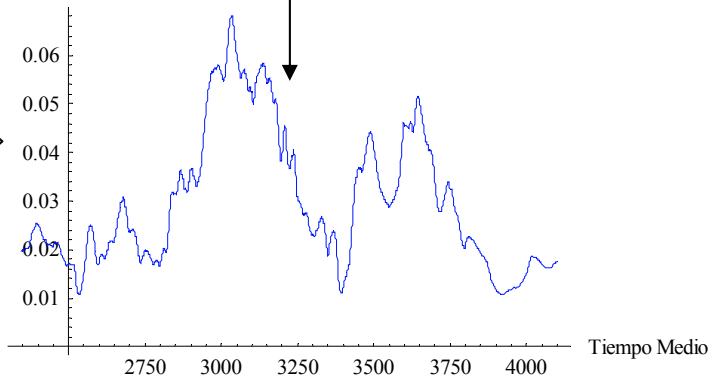
Segundo Gráfico: Conversión de Posición en Tiempo (puente entre los 2 gráficos que suministrar el cambio de escala local).

Tercer Gráfico: Ajuste a la Distribución de Benford en función del tiempo medio en los subconjuntos de corredores.

Tiempo



Ajuste de Distribución



El Equilibrio en Escasez

Comparación

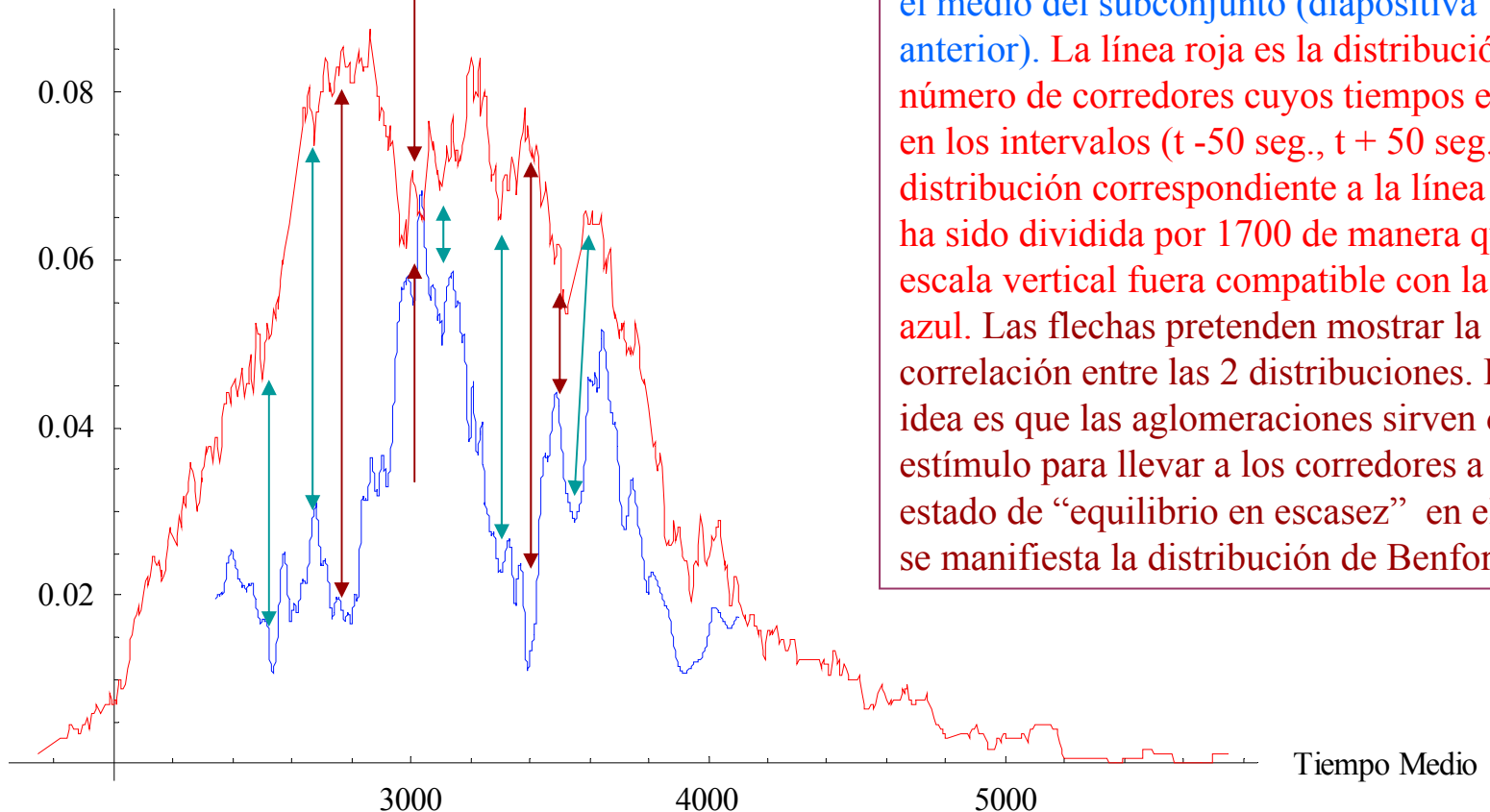
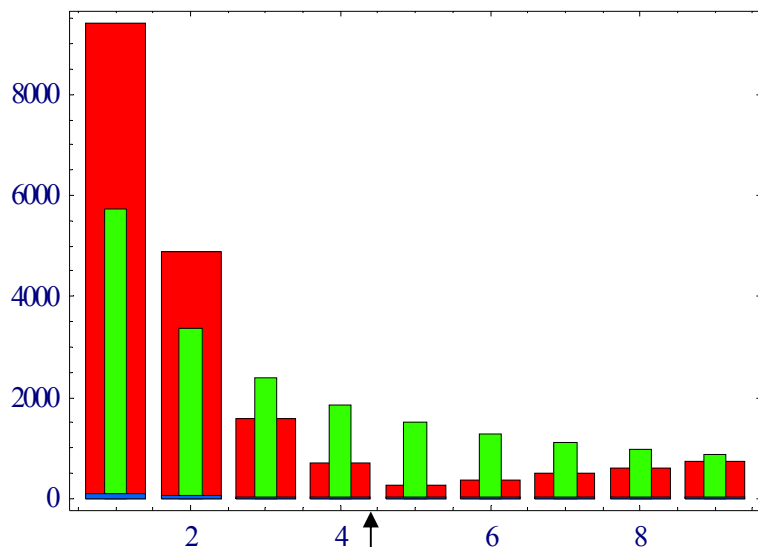


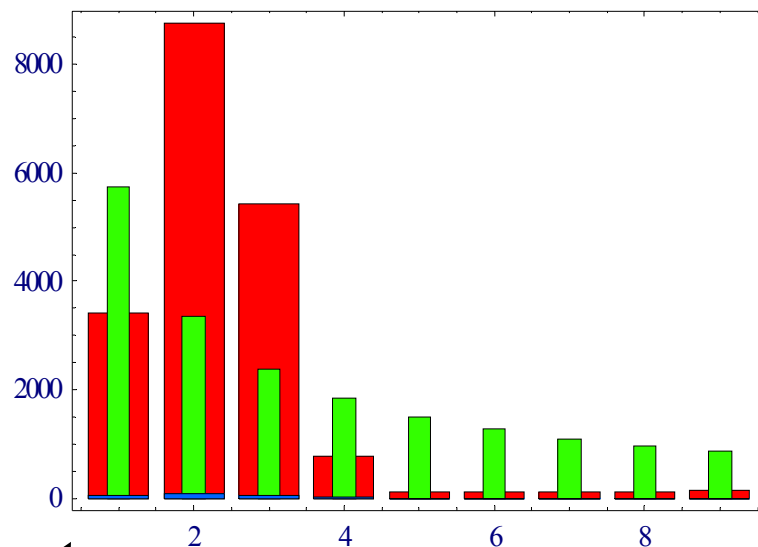
Gráfico Comparativo: La línea azul representa el ajuste móvil de 300 corredores a la distribución de Benford en función del tiempo del corredor ubicado en el medio del subconjunto (diapositiva anterior). La línea roja es la distribución del número de corredores cuyos tiempos están en los intervalos $(t - 50 \text{ seg.}, t + 50 \text{ seg.})$. La distribución correspondiente a la línea roja ha sido dividida por 1700 de manera que su escala vertical fuera compatible con la línea azul. Las flechas pretenden mostrar la anti-correlación entre las 2 distribuciones. La idea es que las aglomeraciones sirven de estímulo para llevar a los corredores a un estado de “equilibrio en escasez” en el cual se manifiesta la distribución de Benford.

Indicios Preliminares en los Datos del Referéndum de Agosto.

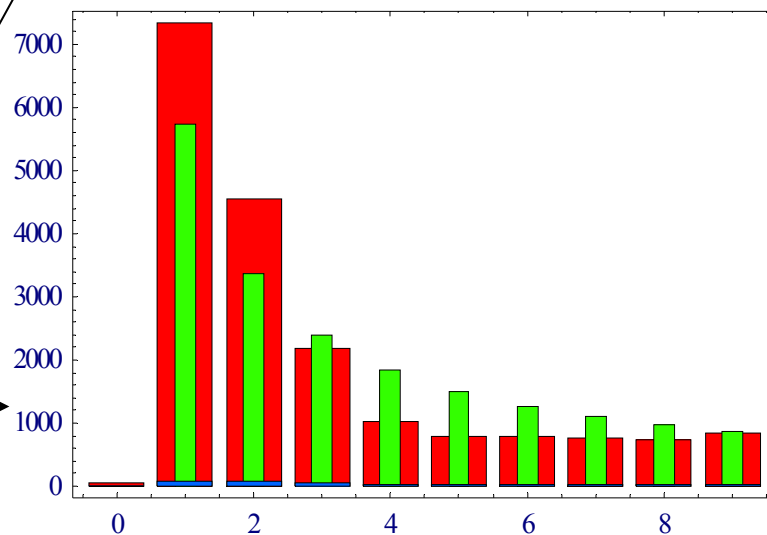
1er Dígito del SI vs. Benford



1er Dígito del NO vs. Benford



1er Dígito de la Resta del NO y el SI vs. Benford



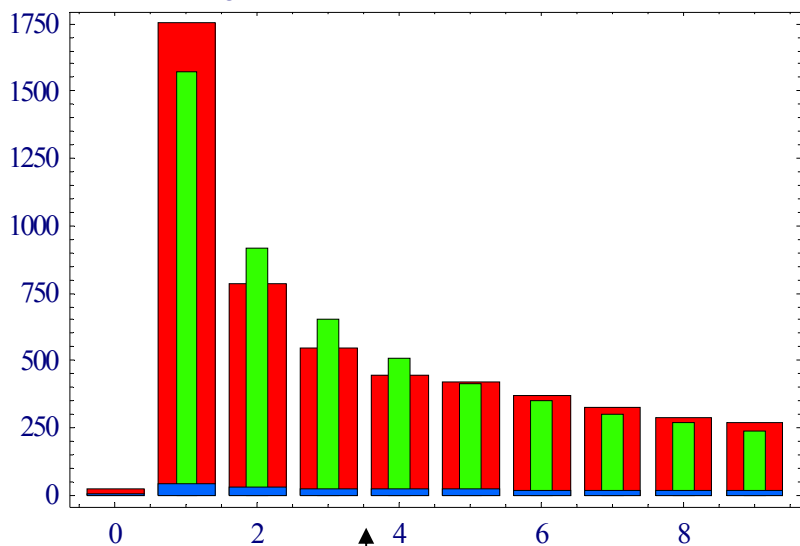
Primer Gráfico: Distribución del primer dígito de los votos SI comparado con la ddB. $S = 0.33$

Segundo Gráfico: Distribución del primer dígito de los votos NO comparado con la ddB. $S = 1.1JK$

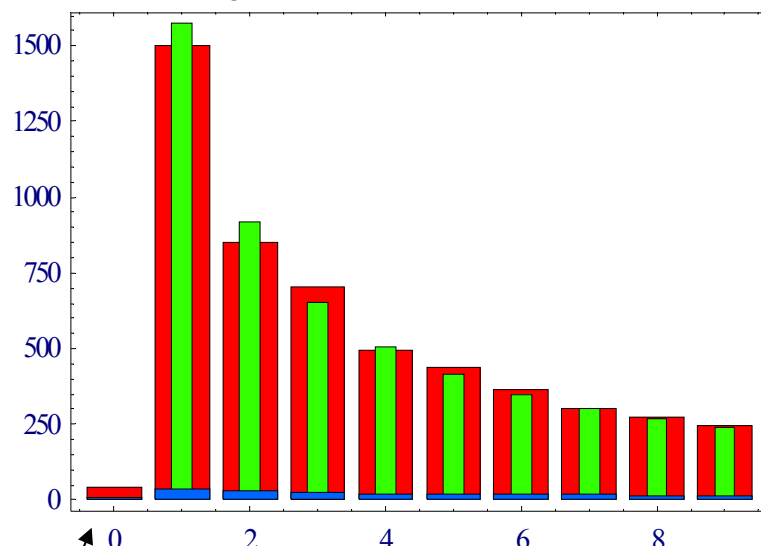
Tercer Gráfico: Distribución del primer dígito de los votos NO-SI comparado con la ddB. $S = 1.1L$

Comparación con las Elecciones del 2.000

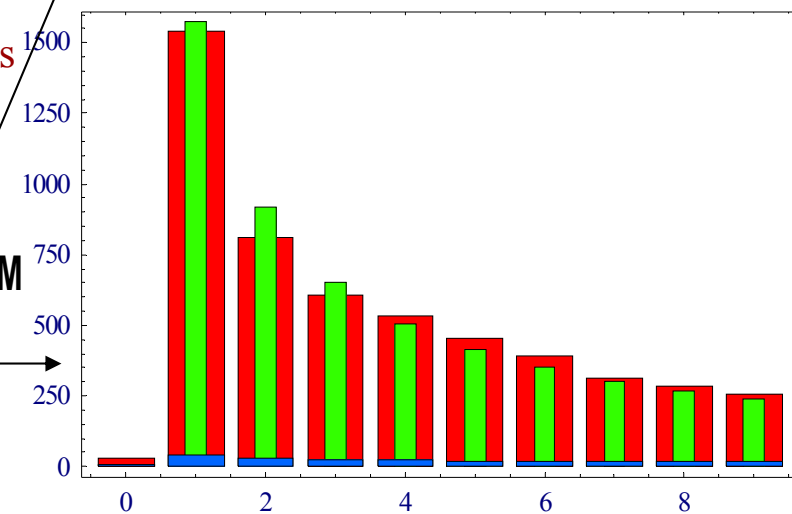
1er Dígito de los Votos a Favor de Chavez vs. Benford



1er Dígito de los Votos a Favor de Arias vs. Benford



1er Dígito de la Resta: Chavez - Arias vs. Benford



Primer Gráfico: Distribución del primer dígito de los votos a favor de Chávez comparado con la ddB. **S = 0.014**

Segundo Gráfico: Distribución del primer dígito de los votos a favor de Arias comparado con la ddB. **S = 0.011**

Tercer Gráfico: Distribución del primer dígito de los votos NO-SI comparado con la ddB. **S = 0.011**

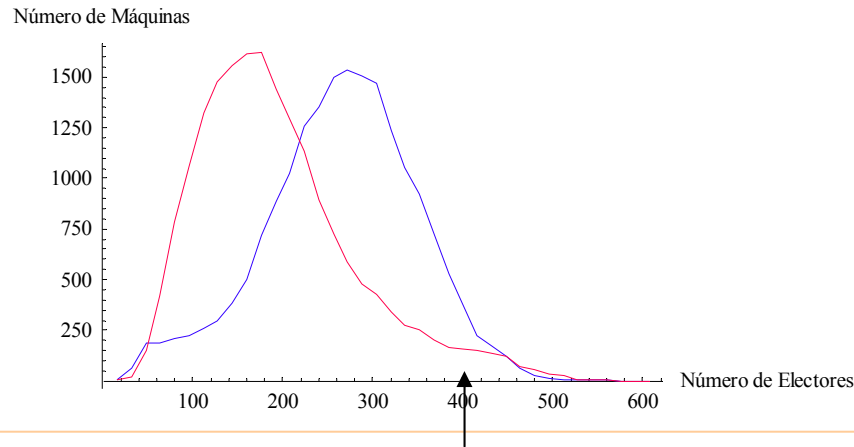
Condiciones Ideales de un Sistema para que Tenga un Buen Ajuste a la Distribución de Benford.

- Ausencia de límites artificiales.
- Condición de Equilibrio en Escasez (hipótesis propia).
- Ausencia de otra distribución que domine los resultados.

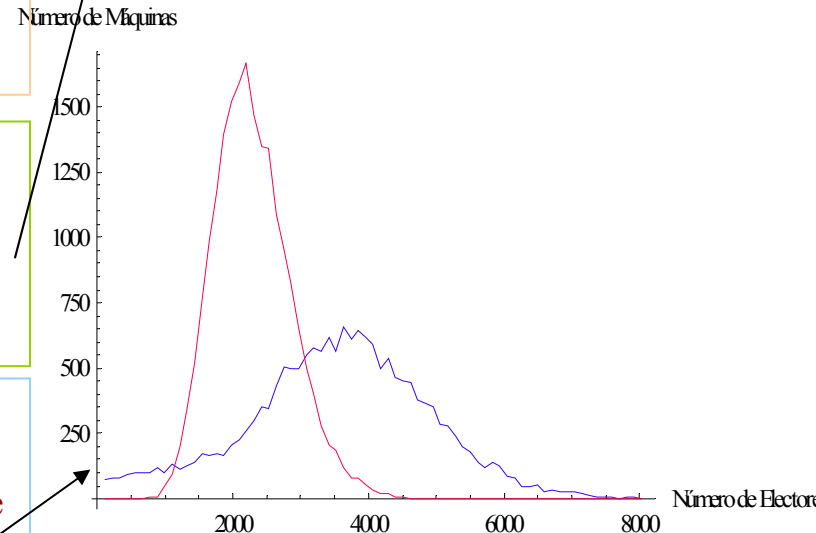
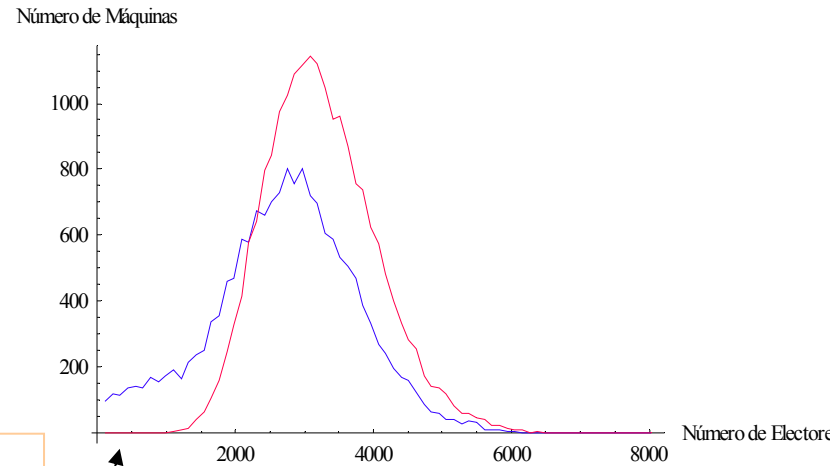
Las hipótesis de este trabajo en relación a los datos del referendo:

- La resta de los votos $\underline{OP} - \underline{\#}$ es un conjunto de números que se ajusta mejor a la distribución de Benford que los conjuntos por separado. Esto se debe a 2 razones; la resta “apantalla” los límites artificiales en los votos por máquina (19.055 máquinas) y la diferencia de votos se aproxima a una condición de equilibrio en escasez ya que las intenciones de voto son muy cercanas al 50-50 (independientemente de quien sea el ganador).
- Bajo la suposición de que la manipulación si existió (se reitera que esta es una hipótesis de trabajo, no una afirmación), al revertirla en función de un parámetro “x”: $(\underline{OP} \cdot (1-x)) - ((\underline{OP} \cdot x) + \underline{\#}) = \underline{OP} - \underline{\#} - (2 \cdot x \cdot \underline{OP})$ y estudiar el ajuste del conjunto resultante a la distribución de Benford, se va a hallar un mínimo que va a estar directamente relacionado con el supuesto porcentaje promedio de votos transferidos artificialmente del conjunto $\underline{\#}$ al conjunto \underline{OP} !

Experimentos Numéricos: Pruebas Empíricas Respecto al Funcionamiento del Método.



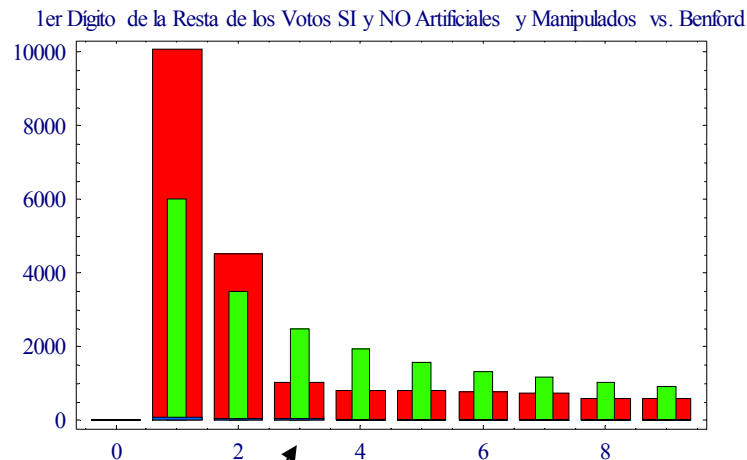
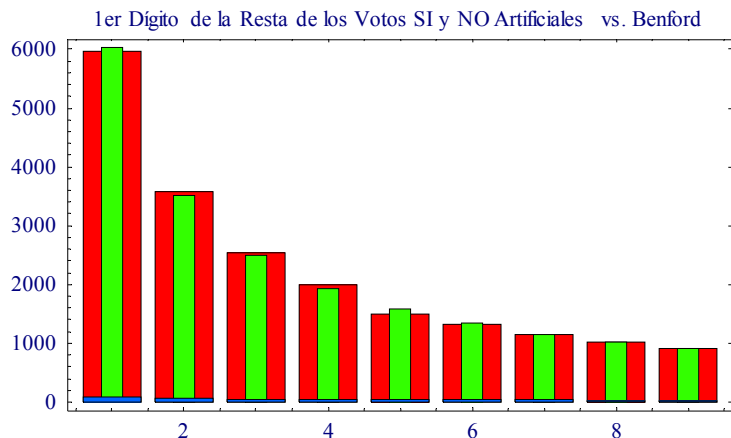
Primer Gráfico: Datos del Referéndum. Distribución del número de máquinas en función del número de votos por máquina en un intervalo de 16 votos. Total de Máquinas: 19.055. Total de Votos Si: 3.584.835. Total de Votos No: 4.917.279



Segundo Gráfico: Datos Artificiales. Distribución del número de máquinas en función del número de votos por máquina en un intervalo de 110 votos. Total de Máquinas: 20.000. Total de Votos Si: 63.742.203. Total de Votos No: 35.152.823. Relación: 65% a 35%. Se usó una distribución binomial negativa.

Tercer Gráfico: Datos Artificiales Modificados. Distribución del número de máquinas en función del número de votos por máquina en un intervalo de 110 votos. Total de Máquinas: 20.000. Total de Votos Si: 44.619.456. Total de Votos No: 54.275.411. Relación: 45% a 55%. Se transfirieron 30% de votos SI a votos NO respecto a los resultados originales (Segundo Gráfico).

Reversión de la Manipulación.

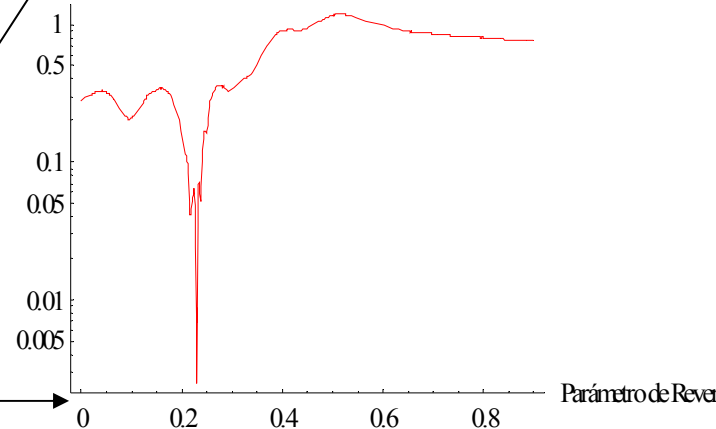


Primer Gráfico: Comparación de la resta de los votos artificiales No y Si, con la ddB. Obviamente el ajuste es casi perfecto porque fueron fabricados expresamente con Montecarlo. $S = 0.0005$

Segundo Gráfico: Comparación de la resta de los votos artificiales No y Si después de haber transferido el 30% de los votos del Si al No, con la ddB. Ahora el ajuste ha empeorado notablemente. $S = 0.28$

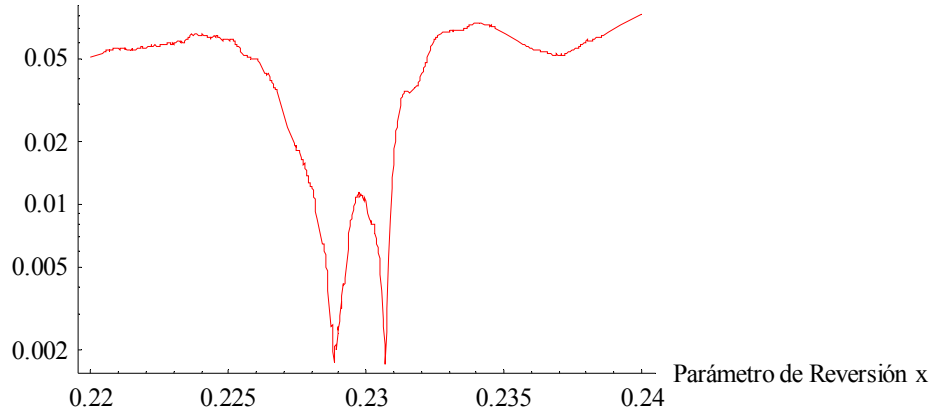
Tercer Gráfico: Ajuste a la ddB en función del parámetro de reversión de los datos artificiales y manipulados. Esta es la primera prueba empírica de que la hipótesis inicial es cierta. Obsérvese que el ajuste mejora en el mínimo en más de **2 ordenes de magnitud** con respecto al resto de la curva!

Ajuste Chi Cuadrado a Benford



Reversión de la Manipulación.

Ajuste Chi Cuadrado a Benford

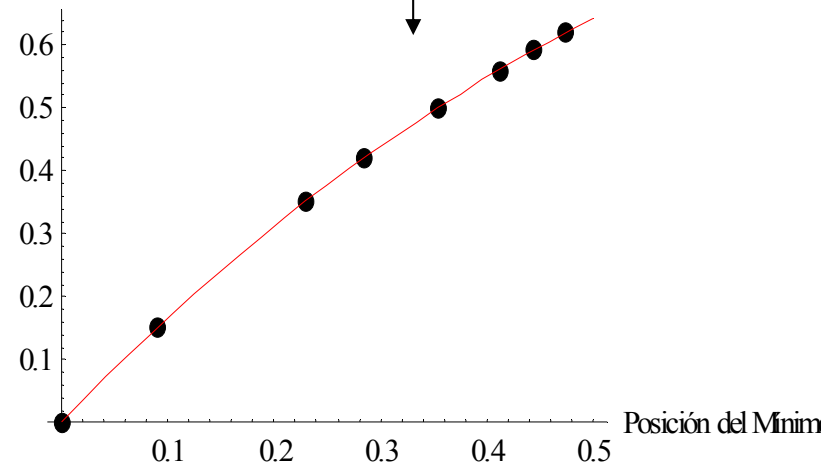


Primer Gráfico: Zoom del mínimo. Obsérvese que hay 2 valles muy cercanos. Se toma la cantidad intermedia de 0.23 como valor único del mínimo. Sin embargo el valor esperado era 0.35. Este valor no es 0.3 debido a que el porcentaje transferido respecto del Si original no es igual al porcentaje respecto del No falseado.

Segundo Gráfico: Existe una dependencia entre la posición del mínimo y el valor real de votos transferidos. Esta dependencia es una parábola muy suave, casi una recta. Esto demuestra que es posible obtener el valor de los votos transferidos a partir de los votos por máquina falseados por medio de la minimización lineal del parámetro de reversión. El ajuste por mínimos cuadrados proporciona la siguiente ecuación:

$$0.00155 C \ 1.7278 x \ H \ 0.89892 x^2$$

Fracción Real Transferida



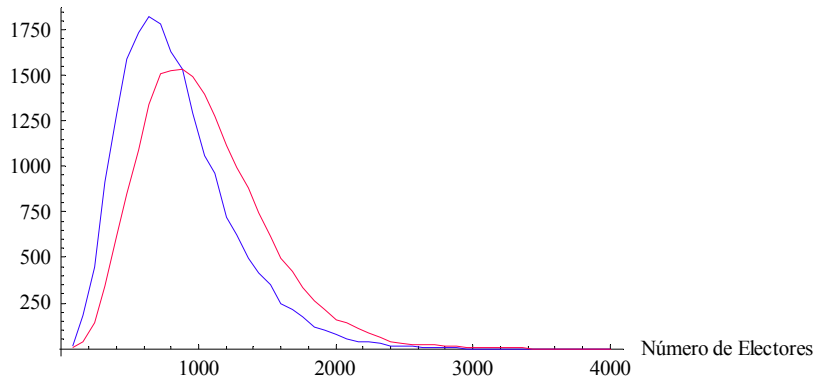
Reversión de la Manipulación.

Hay sin embargo posible críticas a todo el esquema presentado hasta ahora:

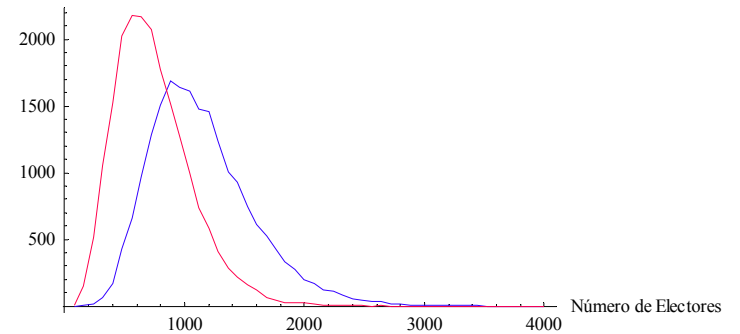
- Si los votos del SI y el NO, no están relacionados en su resta por una distribución que se ajusta bien a la ddB, entonces no habrá mínimos espurios que indiquen una manipulación que nunca ocurrió?
- Es poco probable suponer que la manipulación artificial de los datos halla ocurrido en todas las máquinas. Se podrá encontrar un mínimo en idénticas condiciones a las cuales se hicieron las simulaciones anteriores pero con la diferencia que la transferencia de votos no suceda en todas las máquinas, sólo en una fracción de ellas escogidas por ejemplo aleatoriamente?

Votos del Si y No, no relacionados con ddB

Número de Máquinas

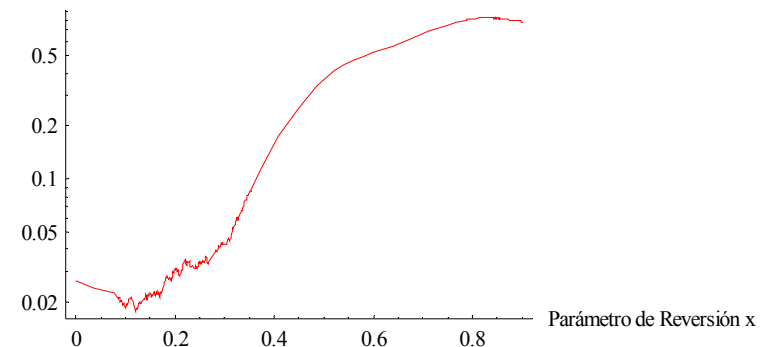


Número de Máquinas

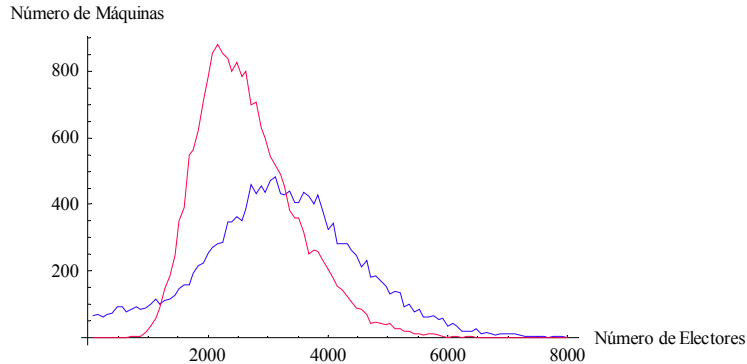


Las dos primeras gráficas muestran un comportamiento similar al caso relacionado por ddB. Sin embargo al graficar el parámetro de reversion, sólo se evidencia un mínimo que es del mismo orden del “ruido” y que además está superpuesto a la tendencia secundaria de la curva.

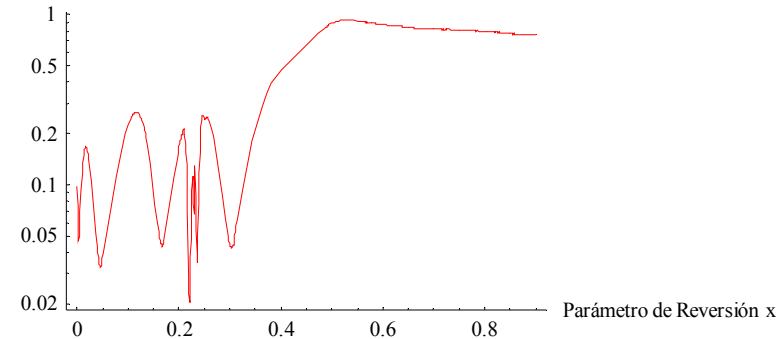
Ajuste Chi Cuadrado a Benford



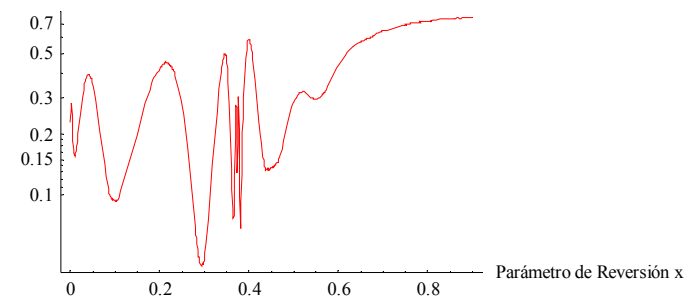
Selección Aleatoria de Máquinas cuyos Resultados van a ser Modificados.



Ajuste Chi Cuadrado a Benford



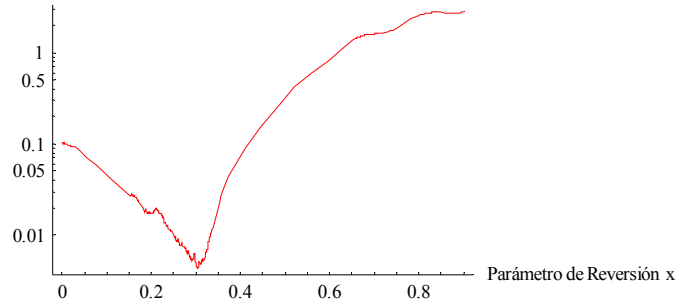
Ajuste Chi Cuadrado a Benford



Si sólo se modifican una fracción de las máquinas, aun es posible detectar un mínimo en el gráfico del ajuste vs. el parámetro de reversión. Pero a medida que la fracción de máquinas afectadas es menor, la profundidad del mínimo va disminuyendo. En el gráfico de la derecha, arriba, se observa el caso en que se modificaron el 60% de la máquinas a con una transferencia del 30% de los votos (lo cual no lograba que ganara el No). La profundidad del mínimo se hace ya comparable a la tendencia subyacente de la línea. Esto significa que se está cerca del límite de detección. El aumentar la transferencia de votos no ayuda, tal como se observa en el gráfico de la derecha, en el cual se transfirieron el 60% de los votos del 60% de las máquinas. Finalmente hay que notar que las distribuciones del número de máquinas en función de los votos, presentan un comportamiento similar al ser modificadas aleatoriamente, sólo un ligero incremento del ruido es observable en la gráfica superior.

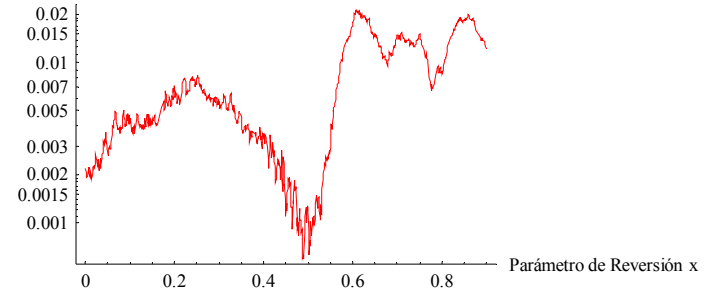
La Minimización del Parámetro de Reversión en los Datos del Referéndum.

Ajuste Chi Cuadrado a Benford



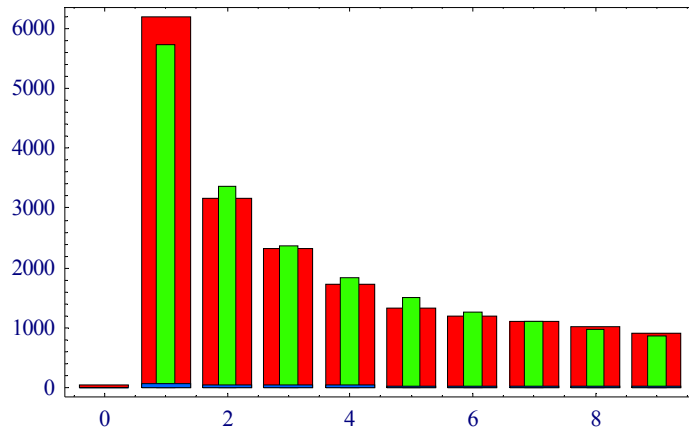
Primer Dígito con 19.055 Máquinas

Ajuste Chi Cuadrado a Benford



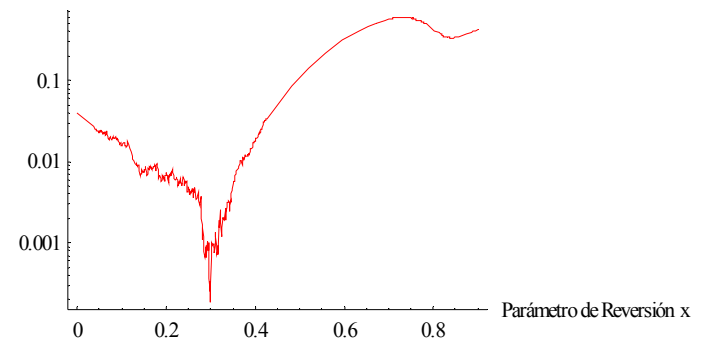
Segundo Dígito con 19.055 Máquinas.

La Distribución en el Mínimo vs. Benford



Comparación en el Mínimo

Ajuste Chi Cuadrado a Benford



Primer Dígito con 7.877 Mesas.

Conclusiones

- El mismo factor que hizo propenso a los resultados del referéndum a ser posiblemente alterados, esto es la cercanía de las tendencias de voto, hace susceptible a la alteración a ser descubierta usando la diferencia de votos y su minimización en cuanto a su ajuste a la distribución de Benford en función del parámetro de reversión, tal como se discutió anteriormente.
- Las simulaciones realizadas dan indicios de la posibilidad de descubrir e incluso cuantificar una alteración en los datos originales de una elección binaria. Usando dichos resultados es posible estimar un promedio de 40% de transferencia de votos (la fracción se refiere a los votos No resultantes después de la alteración).
- Esto último no significa que exactamente el 40% de los votos del O₁ hayan pertenecido originalmente al #*. Las simulaciones también muestran que el método es capaz de detectar alteraciones incluso cuando sólo un porcentaje de los resultados individuales de bien sea las máquinas ó las mesas es modificado. Los mínimos hallados tanto en los datos por máquina como por mesa, presentan signos de provenir de alteraciones en una fracción de las mismas. Siguiendo este orden de ideas y los resultados de las simulaciones se puede estimar una cuota mínima de alteración total de: $0.6 \times 0.4 = 0.24$, esto significa el 24% de los votos del O₁. En números absolutos esto representa aproximadamente 1.390.000 votos.
- El autor finalmente quiere reiterar el carácter científico de este trabajo. La utilidad y efectividad de la Ley de Benford en detectar el fraude en diversos conjuntos de números está bien documentada en numerosos artículos cuyas referencias se van a añadir a este documento próximamente. El autor solo reafirma la validez de sus resultados dentro del marco científico y matemático. Por ningún motivo está ni estuvo en contacto con ningún grupo que hace vida política en Venezuela. Los datos que fueron usados en este trabajo se obtuvieron de páginas Web de libre acceso (incluido el CNE) que también van a ser citadas en corto tiempo.