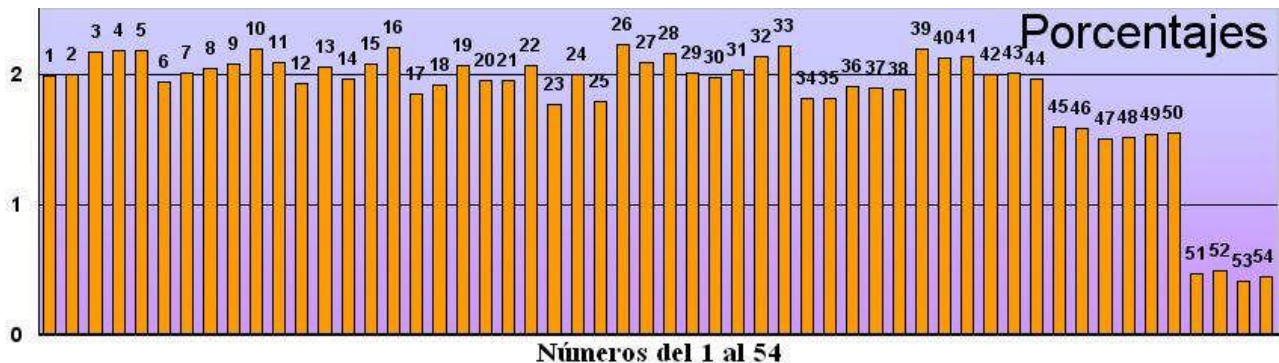


La prueba del último dígito

Jorge A. López, Ph.D.

Una de las pruebas que se usan para determinar si la lotería tiene errores, es un simple conteo de las veces que los números posibles aparecen entre los ganadores. Por ejemplo, la *Lotto* de Texas, que juega con 6 números escogidos entre el 1 y el 54, tuvo la distribución de porcentajes mostrada en la gráfica siguiente entre 1992 y 2006 (www.txlottery.org):



Dado que hay 54 posibles resultados, y suponiendo que todos tienen la misma probabilidad de aparecer, cada número debería tener una probabilidad de $1/54=0.0185185$, o un 1.85185 % de las veces. Aunque la mayoría de los números aparece con porcentajes cercanos a éste, es obvio que conviene apostar preferentemente al 3, 4, 5, 10, 16, 26, 28, 32, 33 y 39, así como también conviene evitar del 45 en adelante, en especial del 51 al 54.

¿Cuál es la causa de esta discrepancia? Sin duda el aparato selector de los números tiene un error sistemático; esto amerita más estudio y una corrección del error.

Un simple conteo es capaz de identificar errores que pasarían desapercibidos.

Lo importante del ejemplo anterior es reconocer que un simple conteo es capaz de ayudarnos a identificar errores que de otra manera pasarían desapercibidos. Este tipo de pruebas: conteo y estudio de frecuencia, son útiles en casos donde hay cantidades grandes de números y donde se espera que estos aparezcan con una cierta frecuencia; tal es el caso de las votaciones.

Este estudio trata de determinar el comportamiento de los datos de la elección presidencial de México del 2006 examinando la frecuencia del último dígito en cada una de las votaciones recibidas por cada partido en cada casilla.

El resultado principal es que el último dígito de los votos tiene una distribución uniforme, excepto para ciertos partidos y en ciertos estados.

Debido a que los argumentos que aquí se presentarán son de indole matemática, posible coco de muchos lectores, se ha hecho un esfuerzo por explicar todos los cálculos de una manera pedagógica incluyendo numerosos ejemplos.

Pero, para aquellos que les da pereza seguir argumentación matemática, les resumo el trabajo en un párrafo. Se analizaron los datos de la elección presidencial, divididos por estados, casillas, y partido. Examinando los votos en cada casilla, se midió la frecuencia con la que aparecen el 0, 1, ..., 9 como últimos dígitos. Como se esperaría, la distribución del último dígito es mayormente uniforme, sin embargo ciertos partidos en ciertos estados tienen distribuciones no uniformes. ¿Qué explica este comportamiento? Como dijo Newton al proponer su ley de gravitación: “No lo sé y no aventuro hipótesis”.

Antecedentes

El análisis de la distribución de dígitos en datos electorales cobró relevancia en años recientes en Venezuela. Analizando el referéndum del 15 de agosto del 2004, Pericchi y Torres (2004) estudiaron la distribución del segundo dígito, y usaron supuestas discrepancias con la distribución esperada (distribución de Benford) para levantar sospechas de fraude. Más tarde, análisis de Brady (2005) y Taylor (2005) fueron usados por el Centro Carter (<http://www.cartercenter.org>) para negar los cargos. [El primer dígito es el más significativo, eg. el “1” en 1567, por ejemplo.]

El análisis de dígitos se ha usado en elecciones en EEUU y en Venezuela

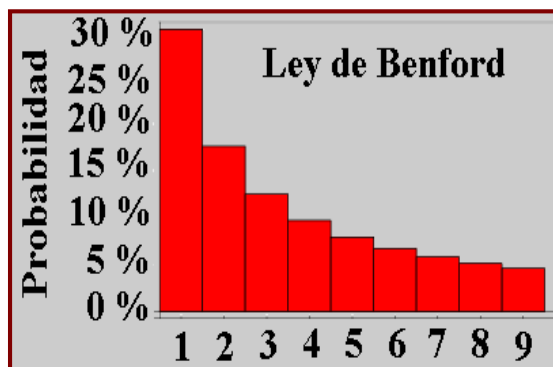
En el 2006, Walter Mebane, profesor de gobierno de la Universidad de Cornell (<http://macht.arts.cornell.edu/wrm1>), también usó la distribución del segundo dígito para estudiar la elección del 2004 en Florida y la del 2006 en México. En sus conclusiones sobre México, el Dr. Mebane afirmó

“... Los resultados por distritos ciertamente sugieren que hay problemas con el conteo de los votos presidencial del 2006 en muchos estados mexicanos, aunque probablemente no en la mayoría. Análisis más refinados se necesitan para llegar a conclusiones más agudas, pero la impresión general es que se amerita una investigación más exhaustiva de la elección.”



Todos estos estudios anteriores se hicieron con la esperanza de identificar desviaciones respecto a la distribución de Benford. Esta ley es una regla *empírica* que afirma que los primeros dígitos de ciertos datos financieros y sociales tienden a seguir una distribución como la mostrada.

En el caso de votaciones, no se espera que el primer dígito siga la distribución de Benford. Por ejemplo, suponiendo que las casillas de algún estado tienen, en promedio, 200 votantes, si se sabe que en ese estado un partido tiene preferencias del 30 %, todas las casillas le reportaran -de nuevo, en promedio— unos 60 votos. Algunos serán de 50s, la mayoría estará en los sesenta y tantos, y habrá unos en los 70s. Obviamente, la cuenta de los primeros dígitos no tendrá su máximo en el “1” como en la distribución de Benford, y es por eso que Mebane, Pericchi y Torres se enfocaron en el segundo dígito.



Pero aún el segundo dígito tiene problemas. Se sabe que muchos datos sociales acatan esta ley, pero se sabe de muchos otros casos, en los que esto no sucede. [El mismo Benford en su estudio inicial (Benford, 1938), usó datos que estadísticamente no satisfacían su propia ley (Scott y Fasli, 2001).] Por lo que demostrar que el segundo dígito no satisface esta ley no significa mucho pues no hay motivos aparentes que indiquen que lo deba de hacer. Por esto, para evitar caer en discusiones sobre si la empírica ley de Benford se aplica o no a los primeros dígitos de datos electorales, nos olvidamos de ellos y nos concentraremos en el último dígito, es decir en el menos significativo.

¿Qué frecuencia deben de tener los últimos dígitos? De acuerdo al IFE, en cada casilla de votación pudo haber un máximo de 760 votos, los cuales fueron divididos entre seis posibilidades (PAN, PRI, PRD, NA, ASDC y nulos). Esta división puede ser hecha en 3,022,160,069 maneras distintas (ver www.btinternet.com/~se16/js/partitions.htm), por lo que existen infinidad de posibilidades de que el último dígito tenga cualquier valor entre el 0 y el 9. En otras palabras, no parece haber restricciones de antemano que favorezcan unos dígitos sobre otros, por lo que es posible suponer que los últimos dígitos de cada votación recibida por cada partido en cada casilla deben tener una distribución uniforme, algo así como lo esperado para la Lotto de Texas, pero limitado del 0 al 9; para este caso se esperaría que cada dígito apareciera en promedio el 10 % del total.

Los datos

El archivo con todos los datos del conteo distrital para la elección de presidente se puede copiar del sitio del Dr. Luis Mochán, ahora director del Instituto de Ciencias Físicas de la UNAM en Cuernavaca:

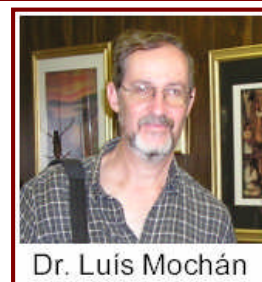
<http://em.fis.unam.mx/~mochan/elecciones/Computos2006-Presidente.zip>



Este archivo fue copiado directamente de la base de datos del IFE:

http://www.ife.org.mx/Computos2006/bd_computos06.htm

que, desgraciadamente, no está más en operación. [¿Porqué?]



Dr. Luis Mochán

Estos datos contienen un poco más de 130,000 líneas con información en el siguiente formato:

```
1|1|0|1|B|0|6|120|0|0|76|76|1|53|4|18|0|1|0||05/07/2006 23:22:10
1|1|338|1|B|0|1|1|689|12|5|341|353|1|202|45|65|12|12|2|1|1|05/07/2006 08:17:05
1|1|338|1|C|0|1|1|689|7|11|291|298|2|175|38|56|3|8|2|1|1|05/07/2006 08:22:24

32|4|1708|1|B|0|1|1|273|1|0|165|166|515|6|75|82|0|2|54|1|1|05/07/2006 18:35:01
32|4|1709|1|B|0|1|1|277|9|1|175|184|516|13|51|108|0|2|54|1|1|05/07/2006 18:35:17
32|4|1710|1|B|0|1|1|485|9|8|183|192|517|23|63|85|2|2|54|1|1|05/07/2006 18:35:36
```

Cada línea corresponde a una casilla y cada número significa, en orden:

ID_ESTADO, DISTRITO, SECCION, ID_CASILLA, TIPO_CASILLA, EXT_CONTIGUA, TIPO_CANDIDATURA, TIPO_ACTA, LISTA_NOMINAL, NO_VOTOS_NULOS, NO_VOTOS_CAN_NREG, NO_VOTOS_VALIDOS, TOTAL_VOTOS, ORDEN, PAN, APM, PBT, NA, ASDC, MUNICIPIO, PAQUETE_ENTREGADO, CASILLA_INSTALADA, FECHA_HORA

Donde los candidatos están identificados como PAN, APM (Alianza por México), PBT (Coalición Por El Bien De Todos), NA (Nueva Alianza) y ASDC (Alianza Social Demócrata Campesina). Cabe notar que veinte casillas contienen datos incompletos del número de votos por lo que tuvieron que ser eliminadas, estas casillas son:

```
2|7|732|1|B, 12|9|348|1|B, 12|9|349|1|B, 13|3|682|1|B, 15|38|4642|1|B,
15|38|4642|1|C, 20|5|197|1|B, 20|5|1472|1|B, 20|5|1472|1|C, 20|5|1472|2|C,
20|6|132|1|B, 20|6|2045|2|E, 20|6|2046|1|B, 20|6|2046|1|C, 20|6|2046|1|E,
20|6|2047|1|E, 20|6|2048|1|E, 20|6|2048|1|B, 20|6|2048|1|C, 20|7|1191|1|E,
```

Los últimos dígitos

El presente estudio se basa en los últimos dígitos. Por ejemplo, para el caso de la siguiente casilla de Zacatecas:

32|4|1708|1|B|0|1|1|273|1|0|165|166|515|6|75|82|0|2|54|1|1|05/07/2006 18:35:01

El número de votos válidos fue de 165, los votos del PAN fueron 6, la APM obtuvo 75, la Alianza PBT ganó 82, la NA no tuvo votos, y la ASDC logró 2 votos. En este caso los últimos dígitos son mostrados en la tabla de la derecha.

Número total de votos: 165		
Partido	Votos	Último dígito
PAN	6	6
APM	75	5
PBT	82	2
NA	0	0
ASDC	2	2

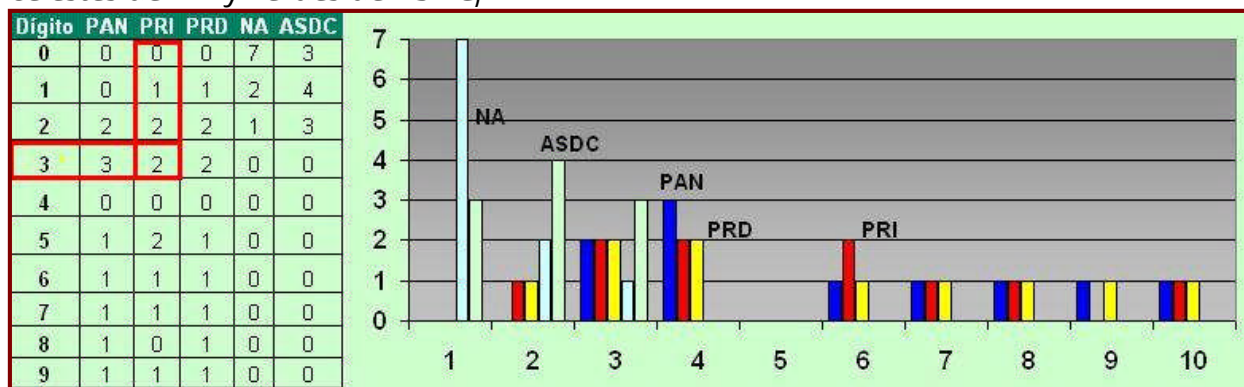
Repitiendo para las últimos diez casillas de Zacatecas

32|4|1701|1|B|0|1|1|256|0|1|157|157|508|32|35|86|1|2|54|1|1|05/07/2006 18:22:00
 32|4|1702|1|B|0|1|1|255|12|0|138|150|509|13|36|82|1|6|54|1|1|05/07/2006 18:24:51
 32|4|1703|1|B|0|1|1|177|3|4|82|85|510|17|13|47|0|1|54|1|1|05/07/2006 18:25:14
 32|4|1704|1|B|0|1|2|119|3|4|71|74|511|25|7|33|0|2|54|1|1|05/07/2006 19:03:57
 32|4|1705|1|B|0|1|1|161|0|2|87|87|512|9|32|43|0|1|54|1|1|05/07/2006 18:33:52
 32|4|1706|1|B|0|1|1|67|1|0|55|56|513|2|32|21|0|0|54|1|1|05/07/2006 18:34:20
 32|4|1707|1|B|0|1|1|181|5|0|96|101|514|18|49|29|0|0|54|1|1|05/07/2006 18:34:44
 32|4|1708|1|B|0|1|1|273|1|0|165|166|515|6|75|82|0|2|54|1|1|05/07/2006 18:35:01
 32|4|1709|1|B|0|1|1|277|9|1|175|184|516|13|51|108|0|2|54|1|1|05/07/2006 18:35:17
 32|4|1710|1|B|0|1|1|485|9|8|183|192|517|23|63|85|2|2|54|1|1|05/07/2006 18:35:36

Se obtiene la siguiente lista de últimos dígitos:

Zacatecas		
Partido	Votos	Últimos dígitos
PAN	32, 13, 17, 25, 9, 2, 18, 6, 13, 23	2, 3, 7, 5, 9, 2, 8, 6, 3, 3
APM	35, 36, 13, 7, 32, 32, 49, 75, 51, 63	5, 6, 3, 7, 2, 2, 9, 5, 1, 3
PBT	86, 82, 47, 33, 43, 21, 29, 82, 108, 85	6, 2, 7, 3, 3, 1, 9, 2, 8, 5
NA	1, 1, 0, 0, 0, 0, 0, 0, 0, 2	1, 1, 0, 0, 0, 0, 0, 0, 0, 2
ASDC	1, 1, 1, 2, 1, 0, 0, 2, 0, 2	1, 1, 1, 2, 1, 0, 0, 2, 0, 2

Y la siguiente distribución (las barras azules son del PAN, rojas del PRI, amarillas del PRD, celestes de NA y verdes de ASDC):



Los números de la tabla bajo cada partido indican las veces que éstos recibieron votaciones que terminaron con el dígito de la primera columna. Por ejemplo, como se indica en rojo, el PRI tuvo 2 votaciones que terminaron en "3": 13 y 63, y su barra de la gráfica correspondiente al dígito "3" sube dos unidades.

Dado que cada partido recibió diez votaciones, en promedio cada dígito debería aparecer $10\% \times 10 = 1$ vez. Aunque NA y ASDC muestran desviaciones grandes respecto a este resultado esperado, los otros tres partidos sí empiezan a acercarse al límite esperado del 10%; esto se vuelve más claro con la inclusión de más casillas.

Las siguientes gráficas muestran la distribución de dígitos en la votación que recibió el PRI en las 1229 casillas en Aguascalientes; la primera gráfica muestra el conteo de dígitos en cantidades, y la segunda en porcentajes.



Esto despeja la duda sobre la distribución que deben seguir los últimos dígitos de los votos recibidos en una elección. Como se puede ver en ambas gráficas, la distribución es muy uniforme y oscila alrededor de los 123 votos, es decir, alrededor del 10%.

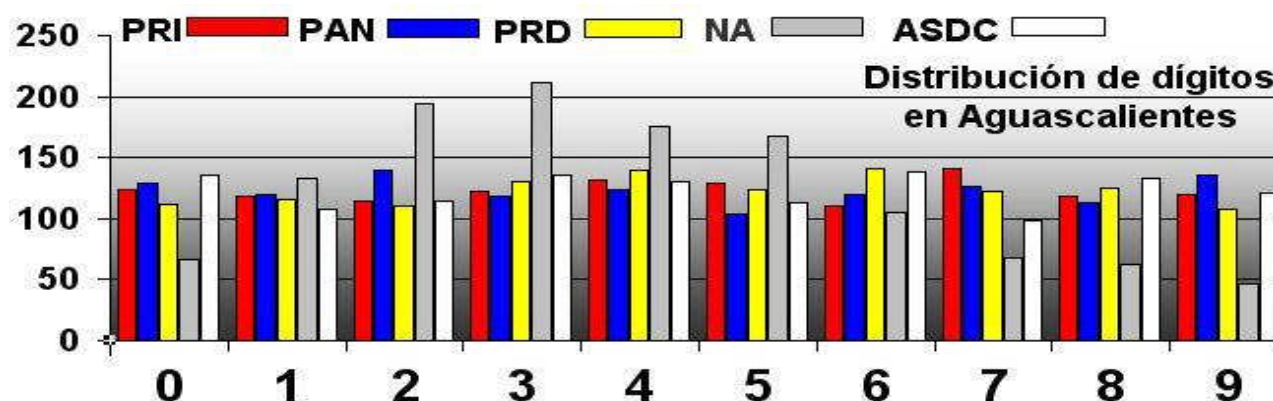
Con esta información, nos abocamos a calcular las distribuciones de dígitos. Para evitar que los resultados de un estado grande, eg. Edo. de México, opaque los de uno chico, eg. Hidalgo, es conveniente hacer estos estudios por estados.

Esto, por otro lado, ayudará a identificar regiones donde pudo haber problemas tales como errores de conteo, etc. El cálculo descrito se puede lograr fácilmente por medio de un programa de computadora; en el presente estudio esto se hizo por medio del programa *Excel* que usa hojas de trabajo y programación en "Visual Basic".

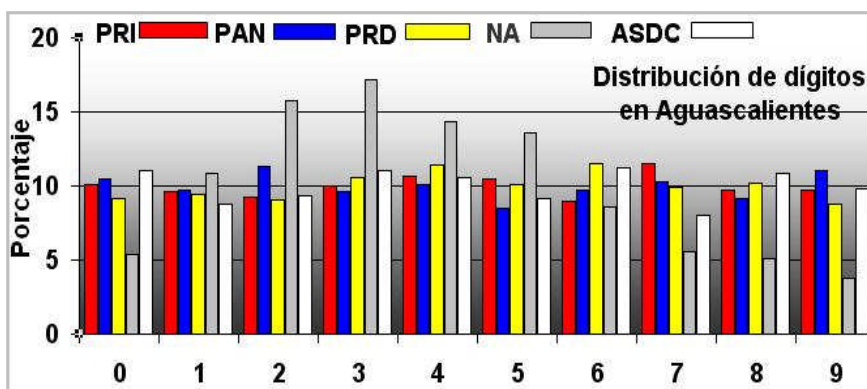


El conteo de dígitos

El método descrito produce la gráfica siguiente para Aguascalientes.



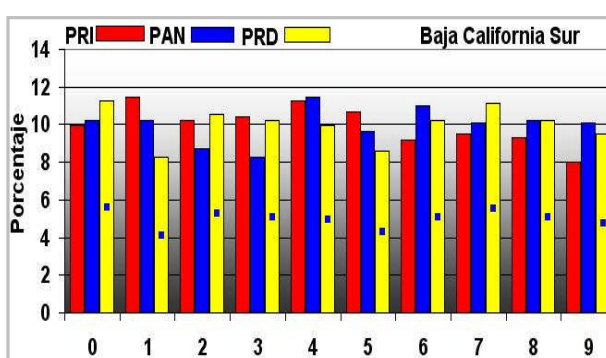
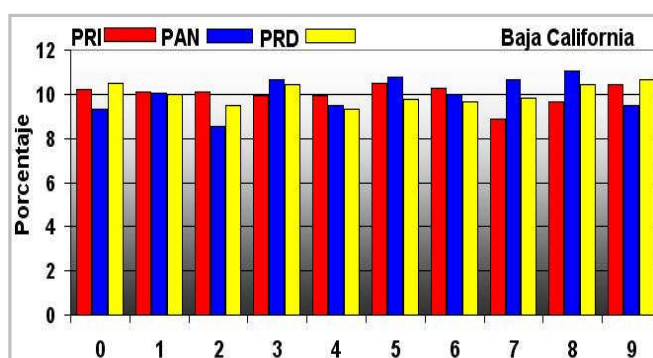
Excepto por los votos de Nueva Alianza (barras grises), los votos de los demás partidos sí parecen ser más uniformes. De nuevo, dado que hubo 1229 casillas, cada partido recibió 1229 cantidades de votos y, por lo mismo, tienen 1229 últimos dígitos. Una distribución uniforme daría alrededor de 123 unidades para cada dígito. La gráfica muestra



oscilaciones cercanas a este valor. En términos de porcentajes la misma información se ve en la gráfica siguiente.

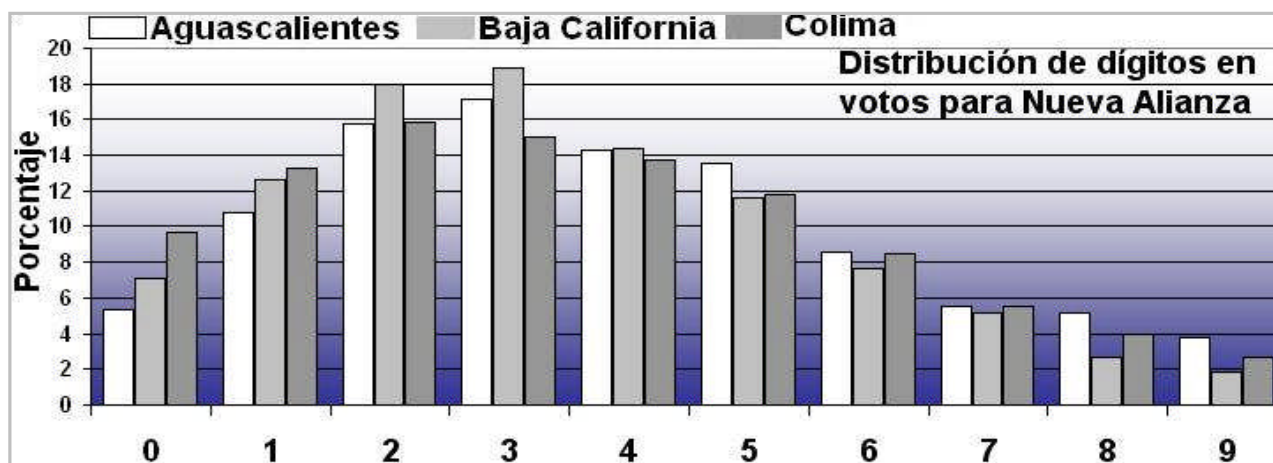
De nuevo, la mayoría de las barras oscilan alrededor del 10 %, con la excepción notable de Nueva Alianza.

Las gráficas siguientes muestran la distribución de dígitos en otros dos estados. Para aumentar la resolución, no se incluyeron los resultados de NA ni de ASDC. Otros estados producen resultados similares, muy cercanos al 10 %.



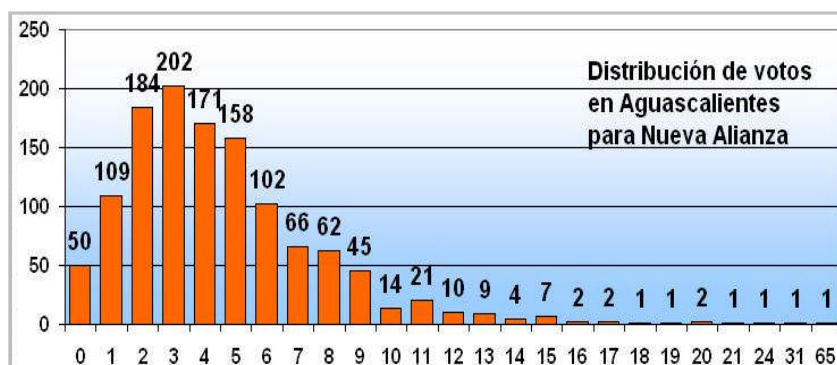
¿Qué pasó con NA y ASDC?

Como se puede ver en la figura antepenúltima, los partidos “chicos”, AN y ASDC, son los que tienen variaciones más grandes respecto a la uniformidad esperada del 10%. Graficando, por ejemplo, la distribución de dígitos de los votos recibidos por NA en Aguascalientes, Baja California y Colima, se puede ver que no siguen una distribución uniforme.



Este resultado es debido a que estos partidos recibieron muy pocos votos –casi siempre menos de 10 por casilla– y esto altera la distribución esperada del último dígito. Cuando un partido grande tiene como último dígito a un “0”, lo más probable es que la votación

recibida haya sido de, digamos, 190 votos o algo substancial terminado en cero. Para AN, sin embargo, un "0" es más comúnmente obtenido con una votación de cero votos.



Esto se puede ver en la gráfica de la izquierda que muestra la distribución de votos para NA en Aguascalientes. Los ceros de la gráfica anterior corresponden a 5.37 % o a 66 ceros. Esta gráfica muestra que 50 de estos ceros corresponden a

votaciones con cero votos, 14 a votaciones con diez votos, y tan sólo 2 a votaciones con veinte votos. Lo mismo sucedió con ASDC, pero a menor escala.

[Dato curioso, con esta distribución de votos, ¿qué tan probable es que NA haya obtenido 65 votos en una casilla, como se muestra en la última columna?]

¿Qué tan creíble es que NA reciba 65 votos en una casilla cuando su promedio fue de 3 por casilla?

Cerramos esta sección con tres observaciones:

- ⊙ Partidos con votación abundante tienen una distribución uniforme de últimos dígitos.
- ⊙ Partidos con poca votación no siguen una distribución uniforme de últimos dígitos.
- ⊙ La distribución de Benford no se aplica en el caso de últimos dígitos.

Midiendo la uniformidad

El "dato curioso" mencionado en la última sección nos hace considerar si es creíble que NA haya recibido 65 votos en una casilla en un estado donde, en promedio, recibió tan solo 3 por casilla. Afortunadamente, esta pregunta ha sido formulada y respondida en el contexto estadístico.

En general, el problema se reduce a medir la desviación de los datos respecto a la distribución normal. Tomemos como ejemplo la votación recibida por el PAN en Chiapas.



Con 3094 casillas, cada partido recibió igual número de votaciones, una distribución uniforme de últimos dígitos resultaría en, aproximadamente, 310 ceros, 310 unos, etc. La gráfica siguiente muestra que los resultados del PAN distan mucho de esta uniformidad esperada (línea roja).

Es obvio que, entre otras desviaciones,, esta gráfica tiene un exceso de votos terminados en "6", más del 11% para ser más precisos. ¿Qué significa este exceso? ¿Deberemos de gritar ¡fraude! o es una fluctuación de las que siempre existen?

Chi cuadrado

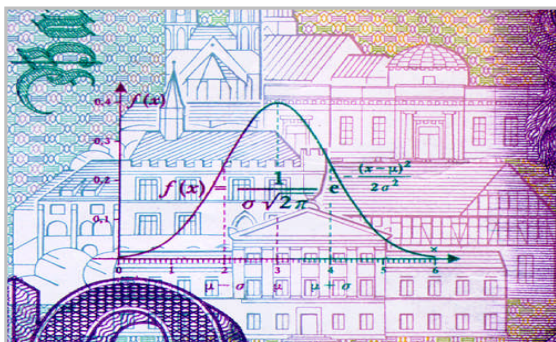
Siguiendo con el caso anterior, y si una desviación del 11 % es aceptable, ¿lo será una del 35 %? ¿O del 60 %? Eso nos lleva a la pregunta: ¿cómo decidir si un resultado es aceptable o no? La respuesta nos la da la prueba del *Chi cuadrado*, χ^2 , que no tiene nada que ver con Tai Chi ni con ninguna otra rama de la filosofía, aunque algo tiene de filosófico.

El chi cuadrado fue estudiado inicialmente por el matemático inglés Karl Pearson quien lo propuso como prueba para medir significados estadísticos. La parte filosófica no viene de Pearson –quien renunció al cristianismo y cambió su nombre de Carl a Karl en honor a Marx– sino que viene del hecho de tener que decidir, por ejemplo, que tanto está uno dispuesto a aceptar que le ganen con dados *cargados* antes de reclamar.

Los dados deben, en promedio, dar la misma cantidad de caras en una serie de tiros, si una cara sale más a menudo que las demás, ¿debe uno reclamar? Claro que la respuesta debe ser sí, pero entonces la pregunta a responder ahora es: ¿qué tan a menudo debe salir una cara más que las demás para que el reclamo sea justo? ¿Un 50 % más que las demás?, ¿un 10%?, un 1 %? La respuesta nos la da una comparación con un dado bien hecho.

Todos los procesos que se miden muestran variaciones de medición a medición. Para saber si una serie de mediciones corresponde *estadísticamente* con otra serie anterior se toma la primera serie como correcta y se determina la probabilidad que la nueva serie se pueda obtener de un *muestreo* de la primera.

Un muestreo, estrictamente hablando, consistiría de un número muy grande de tiros con un dado bien hecho que permitiera saber que tan probable es que se lleguen a observar fluctuaciones del 50 %, 10 % o 1%. En el caso de excesos del 1% de una cara sobre las demás, es posible que el muestreo indique que estos se dan en todas las series de tiros suficientemente largas, por lo que en ese caso no habría nada que discutir. Pero si en el caso de variaciones de 50 %, el muestreo arroja que no es posible lograr variaciones tan grandes en series de tiros razonablemente largas, un reclamo en este caso estaría justificado.



esperados— siguen una distribución normal, la cual nos permite determinar la probabilidad de su existencia. La distribución normal, que ha sido usada para describir infinidad de mediciones, apareció en la última edición de billetes de 10 marcos alemanes en honor a Gauss.



REAL ACADEMIA ESPAÑOLA

DICCIONARIO DE LA LENGUA ESPAÑOLA

Escriba la palabra que desea consultar:

muestreo.

1. m. Acción de escoger muestras representativas de la calidad o condiciones medias de un todo.
2. m. Técnica empleada para esta selección.
3. m. Selección de una pequeña parte estadísticamente determinada, utilizada para inferir el valor de una o varias características del conjunto.

En la práctica, el muestreo no es necesario para determinar la probabilidad de existencia de una fluctuación dada; esta probabilidad se calcula por medio de la teoría de errores. La ley del error, de Laplace, y Gauss del siglo 19, dice que los errores – es decir, las diferencias entre mediciones y los valores



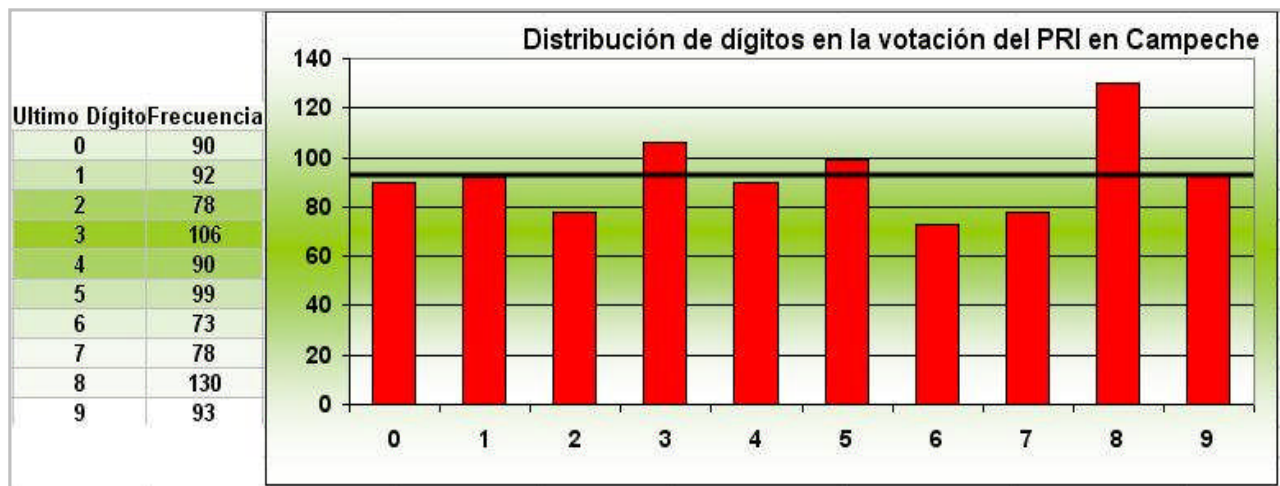
Karl Pearson (1857-1936)

Para usar esta teoría, el “error”, o más apropiadamente, el chi cuadrado, que corresponde a una diferencia promedio normalizada entre una serie de N mediciones, $\{O_1, O_2, \dots, O_N\}$, y la serie esperada, $\{E_1, E_2, \dots, E_N\}$, se calcula de la siguiente manera:

$$\chi^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}.$$

Una vez calculado el χ^2 para la serie, se ubica su valor en la curva normal y se determina la probabilidad de existencia. En nuestro caso, es posible hacer todos estos cálculos y comparaciones de una manera automatizada usando las funciones *CHITEST*, que calcula el χ^2 , y el *CHIINV*, que obtiene la probabilidad de existencia de ese valor de χ^2 , ambas funciones de *Excel* y *Visual Basic*.

El significado de todo esto será más claro con un ejemplo. La gráfica siguiente muestra la distribución de últimos dígitos obtenida por el PRI en las 929 casillas de Campeche.



Una distribución uniforme tendría, aproximadamente, 93 ceros, 93 unos, etc., representado en la gráfica con la línea negra. Para saber que tan probable es que la distribución de dígitos (barras rojas) pueda ser producida por medio de un muestreo aleatorio de la distribución uniforme (línea negra), se calcula χ^2 usando la frecuencia como la serie observada: $\{O\} = \{90, 92, \dots, 93\}$ y el promedio como la serie uniforme esperada: $\{E\} = \{92.9, 92.9, \dots, 92.9\}$.

Con estos datos, la chi cuadrada nos da

$$\chi^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i} = \frac{(90 - 92.9)^2}{92.9} + \frac{(92 - 92.9)^2}{92.9} + \dots + \frac{(93 - 92.9)^2}{92.9} = 26.296.$$

Valor que, una curva normal producida (como en este caso) con 10 números (o 9 *grados de libertad*) nos indica que tiene una probabilidad del 0.0018270 de ser producida en un muestreo aleatorio de la serie esperada. Es decir, el resultado observado se produce, en promedio, una vez cada $1/0.0018270 \approx 547$ muestreos; un resultado poco común, en realidad.

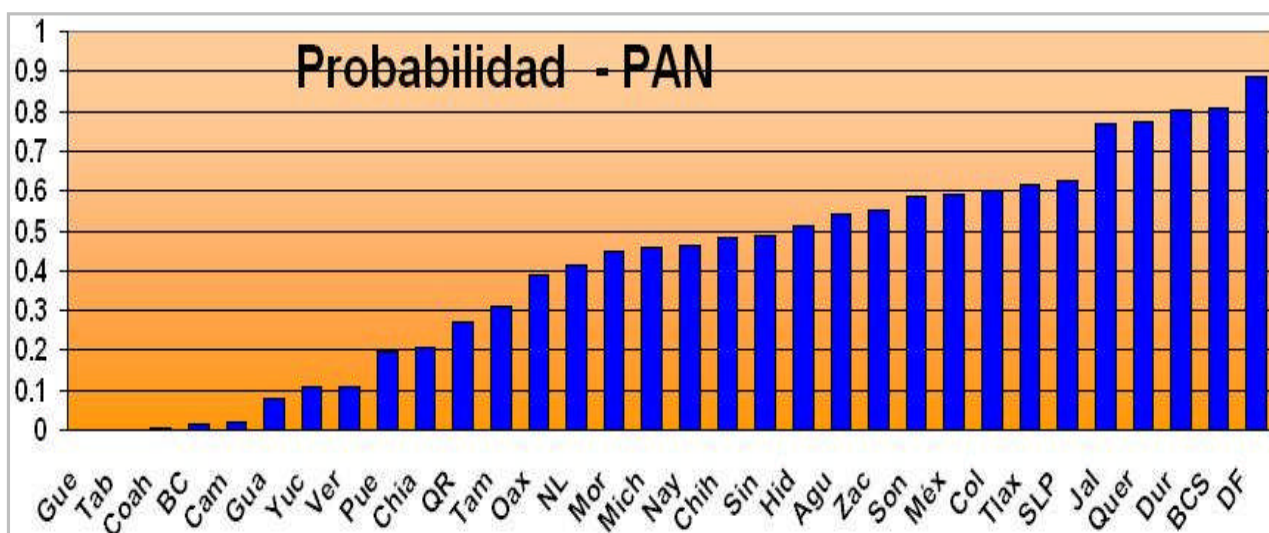
Es claro que el problema viene del exceso de ochos. Si en lugar de tener 130 ochos, tuviéramos, digamos, 93, la chi cuadrada sería $\chi^2 = 11.480$ y existiría una probabilidad de 0.244239, es decir, sería un valor del chi cuadrado que aparecería en promedio una vez cada 4 series; algo muy frecuente, en realidad.

El número excesivo de votos terminados en "8" recibidos por el PRI en Campeche, se puede dar tan sólo una vez en 574 elecciones

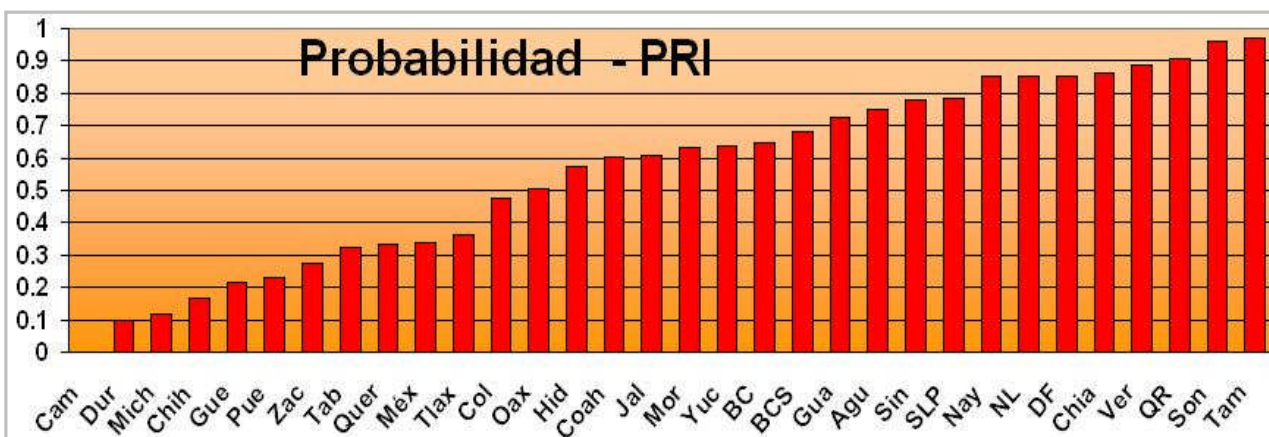
Normalmente, la probabilidad que se usa como límite razonable, *p-value* como es llamado en la jerga estadística, es del 10 %. Valores de chi cuadrada que resultan en probabilidades menores al *p-value* adoptado se consideran muy poco frecuentes, y las series que los producen se toman como distintas a la esperada.

Chi cuadrado de los datos de la votación

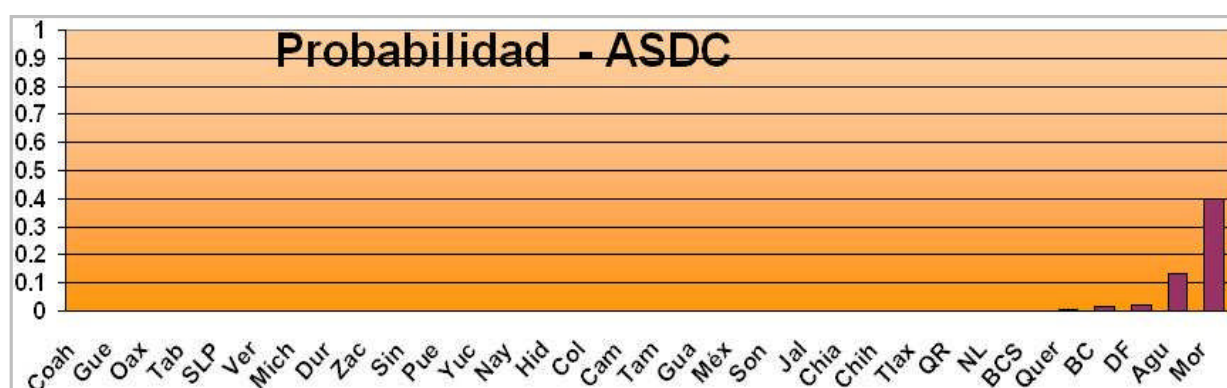
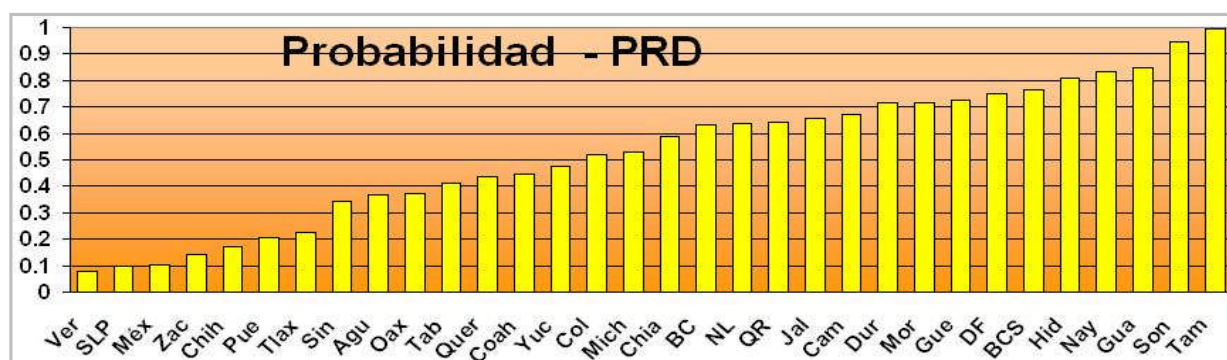
Si aplicamos el método anterior para calcular las chi cuadradas y sus probabilidades correspondientes se obtienen los siguientes resultados para el PAN en los 32 estados, ordenados de menor a mayor probabilidad.



Resultados inmediatos son que la mayoría de los estados satisfacen la regla de uniformidad de la distribución del último dígito en las votaciones recibidas en cada casilla, y seis estados, Guerrero, Tabasco, Coahuila, Baja California, Campeche y Guanajuato (Yucatán tiene probabilidad de 0.1090) tienen probabilidades menores al 10 % de que sus distribuciones de últimos dígitos sean considerados como uniformes.



Los demás partidos dan los siguientes resultados. El PRI no cumple con la regla de uniformidad tan solo en Campeche y Durango; el PRD falla en Veracruz y en San Luis Potosí.



Como se explicó anteriormente, la NA y la ASDC no cumplen con la regla de uniformidad. Haciendo el estudio para la ASDC nos da la última gráfica. La de la NA sería una gráfica vacía con barras de altura cero.

Resumen y conclusión

Se analizaron los datos de la elección presidencial, divididos por estados, casillas, y partidos. Examinando los votos en cada casilla, se midió la frecuencia con la que aparecen el 0, 1, ..., 9 como últimos dígitos en las cantidades de votos recibidas. Como se esperaba, la distribución del último dígito es mayormente uniforme, es decir todos aparecen con frecuencias muy similares. Sin embargo votaciones para el PAN en seis estados (Guerrero, Tabasco, Coahuila, Baja California, Campeche y Guanajuato), para el PRI en dos (Campeche y Durango) y para el PRD en dos (Veracruz y San Luis), no tienen distribuciones uniformes y corresponden –sobre todo las de Guerrero, Tabasco, Coahuila, Baja California y Campeche– a resultados irrealizables en la práctica.

Los resultados de Guerrero, Tabasco, Coahuila, Baja California y Campeche, corresponden a resultados irrealizables

¿Qué tan graves son las irregularidades encontradas en este estudio?

La tabla siguiente pone estos resultados en perspectiva. Los resultados que obtuvo el PAN en Guerrero son repetible una vez cada 669,609 elecciones, los de Tabasco una vez cada 1024 elecciones, etc. Por otro lado, los de Puebla del PRD se podrían esperar una vez cada seis elecciones

PAN		PRI		PRD	
669608.5	Gue	547.327	Cam	12.39303	Ver
1023.872	Tab	10.09579	Dur	10.12608	SLP
334.0617	Coah	8.389657	Mich	9.799927	Méx
65.76256	BC	6.038724	Chih	7.081876	Zac
47.26676	Cam	4.682498	Gue	5.908496	Chih
12.63832	Gua	4.357569	Pue	4.842283	Pue

¿Qué explica este comportamiento? Como dijo Newton al proponer su ley de gravitación:

“No lo sé y no aventuro hipótesis”.

Referencias

- © Benford, F., 1938, The Law of Anomalous Numbers. *Proc American Philosophical Society*, 78(4), pp 551-772.
- © Brady, Henry E. 2005. “Comments on Benford’s Law and the Venezuelan Election,” Enero 19, 2005.
- © Carter Center, “Observing the Venezuela Presidential Recall Referendum: Comprehensive Report,” 2005
- © Walter R. Mebane, Jr., “Election Forensics: Vote Counts and Benford’s Law,” presentado en la reunión de verano del 2006 de la Sociedad de Metodología Política, Universidad de California - Davis, Julio 18, 2006
- © Pericchi, Luis Raul y David Torres, “La Ley de Newcomb-Benford y sus aplicaciones al Referendum Revocatorio en Venezuela,” Reporte Técnico no-definitivo 2a. versión: Octubre 1, 2004.
- © Scott, P. D. y M. Fasli, “Benford’s Law: An Empirical Investigation and a Novel Explanation”, CSM Technical Report 349, Department of Computer Science, University of Essex, 2001.
- © Taylor, Jonathan, “Too Many Ties? An Empirical Analysis of the Venezuelan Referendum Counts,” November 7, 2005.