

# Comprobación de la hipótesis de Delfino y Salas

por Rodrigo Medina A.  
Doctor en Física, [rodrmedina@gmail.com](mailto:rodrmedina@gmail.com)

## Resumen

*Se comprueba, en acuerdo con la hipótesis recientemente publicada por Delfino y Salas, que los resultados oficiales del Referendo Revocatorio de 2004 en cada centro automatizado fueron inventados a partir del número de firmas recogidas en ese centro para pedir el referendo. Se propone una posible forma del algoritmo usado. Se demuestra que el conjunto de centros que terminó siendo auditado en caliente está correlacionado con el número de firmas que se recogieron en esos centros. Se muestra que las anomalías de una encuesta a salida de urna sólo son explicables si los resultados del revocatorio fueron manipulados de alguna manera relacionada con las firmas. Se estimó la magnitud de la diferencia entre el resultado oficial y el verdadero en 1.370.000 votos con un error del 20%*

7 de marzo de 2006

## 1. La hipótesis de Delfino y Salas.

G. Delfino y G. Salas [1] han dado a conocer recientemente un estudio del Referendo Revocatorio Presidencial de 2004. Se puede encontrar allí una descripción bastante detallada del referendo y las referencias a los documentos relevantes. En ese trabajo los autores citados hacen la siguiente hipótesis (la traducción libre del inglés es mía):

*“En los centros automatizados, los resultados oficiales fueron forzados de manera de que los votos por el SÍ siguieran una relación lineal con el número firmas recogidas en el centro para pedir el referendo.”*

Delfino y Salas basan su hipótesis en la gran diferencia que se observa entre los centros automatizados y los centros manuales cuando se estudia la relación entre votos y firmas. En particular si observamos el gráfico de votos por el SÍ contra el número de firmas recogidas en cada centro (Fig. 1), vemos que en los centros automatizados hay una correlación excepcionalmente buena, (coeficiente de correlación  $r = 0.989$ ). Un coeficiente de correlación  $r = 1$  indica una correlación perfecta, o sea un gráfico en donde todos los puntos caen en una línea recta inclinada. Ese resultado fue usado por el Centro Carter para argumentar que el resultado del RR era correcto.

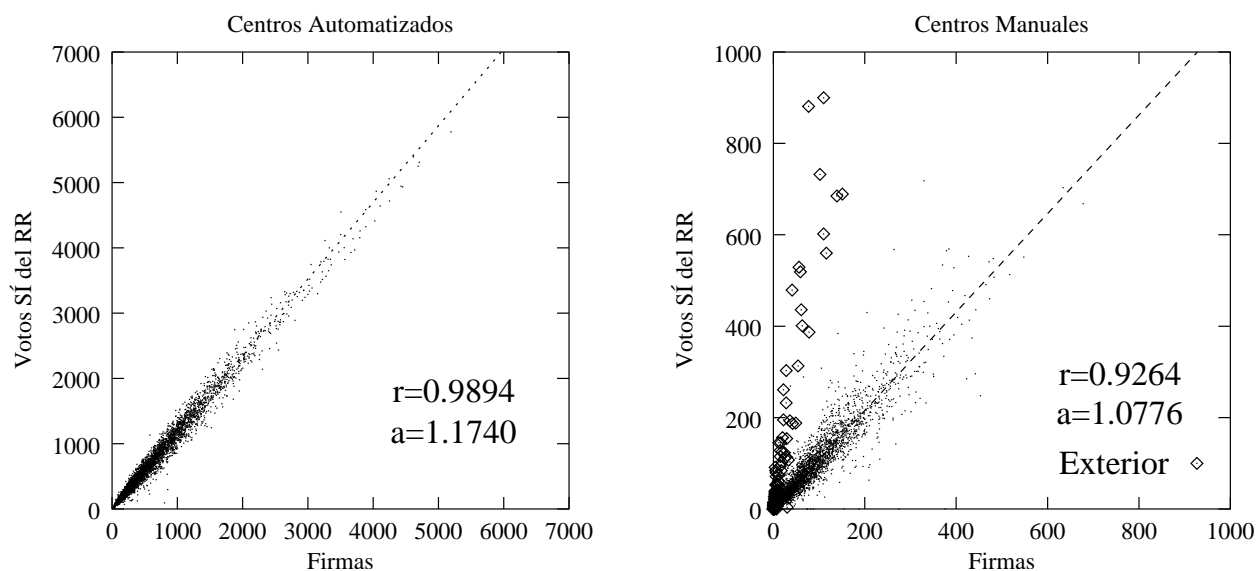


Fig. 1 Votos *vs* firmas para los centros manuales y automatizados. Los centros del exterior se comportan de manera tan diferente del resto de los manuales que fueron excluidos en el cálculo de la correlación. Los coeficientes de proporcionalidad  $a$  se calcularon simplemente dividiendo el total de votos entre el total de firmas.

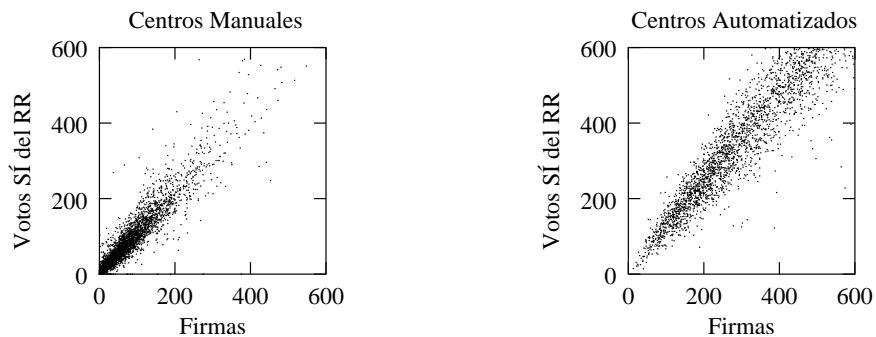


Fig. 2 Comparación para números pequeños

Por otra parte la correlación es bastante peor en los centros manuales, ( $r = 0.9264$ ) que en los automatizados. La diferencia de comportamiento es mayor en la zona de bajo número de firmas y votos, en donde es evidente que los centros manuales presentan una dispersión mucho mayor de los resultados (ver Fig. 2). No hay ninguna razón para que centros del mismo tamaño, muchos de los cuales tienen las mismas características socio-económicas se comporten tan diferentemente por el hecho de estar o no automatizados.

Los autores citados encontraron una manera de hacer más evidente las diferencias. Ellos definen dos variables  $f$  y  $k$ , (en realidad ellos en vez de  $f$  usan  $s$ , por *signature*):

$$f = \frac{\text{Firmas}}{\text{Votos Totales}}$$

$$k = \frac{\text{Votos SI}}{\text{Votos Totales}} = \frac{\text{Votos SI}}{\text{Firmas}}$$

En la Fig. 3 se encuentra graficada la variable  $k$  contra  $f$  para los centros manuales y automatizados. Cada punto representa un centro. Encontramos una diferencia evidente entre manuales y automatizados.

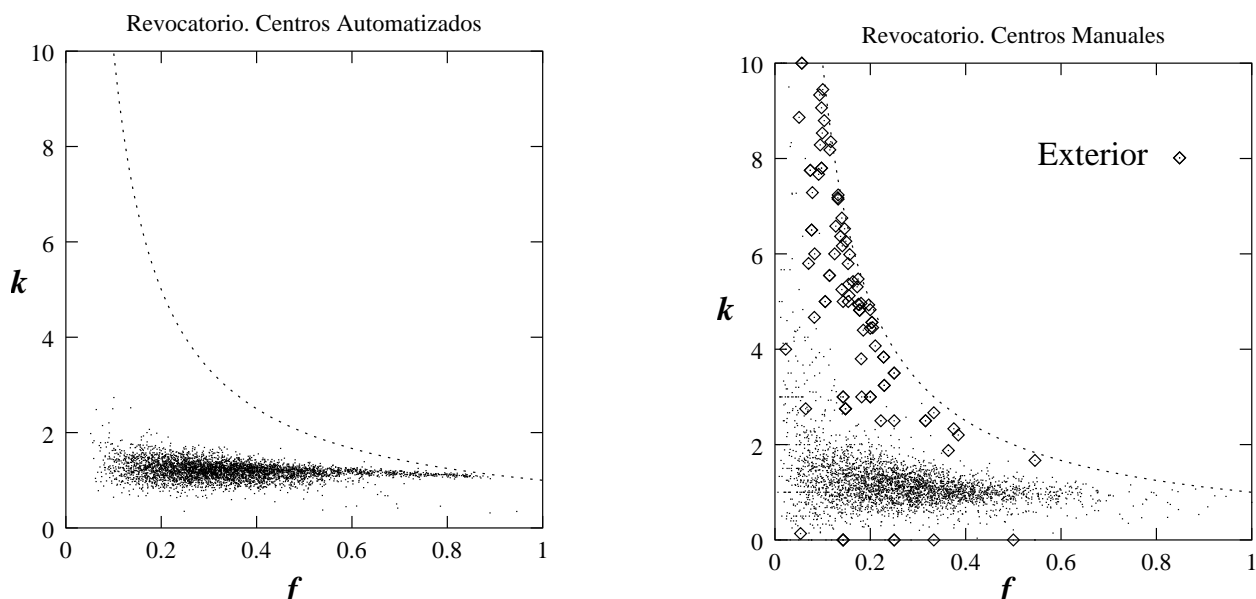


Fig. 3  $k$  contra  $f$  para el RR. La línea punteada corresponde al máximo valor de  $k$ , cuando todos los votos son SI

En los automatizados casi no hay puntos por encima de  $k = 2$  (7 de 4588 y ninguno llega a  $k = 3$ ), mientras que en los manuales hay una nube de puntos por encima y la dispersión es mucho mayor. Cabría esperar una gran dispersión de puntos en ambos casos, manuales y automatizados, debido a la diferencia que había entre firmar y votar. El caso extremo fueron los centros del exterior donde no se podía firmar. Fueron varios los factores que impidieron que muchas personas que querían firmar pudieran hacerlo, además esa dificultad de firmar estuvo distribuida muy desigualmente entre los centros. Entre esos factores podemos citar: a) el número de planillas para las firmas cubría sólo una fracción del electorado (aproximadamente el 30%), b) las planillas estaban distribuidas uniformemente en el país, independientemente de la distribución geográfica de los potenciales firmantes, c) el número de centros de recolecta de firmas era sólo un tercio del número de centros de votación, esto hizo que en zonas rurales los centros de recolección estuvieran lejos de muchos centros poblados, d) no es lo mismo asistir a una elección secreta a un centro donde todos asisten, que identificarse como opositor por el mero hecho de asistir a un centro de recolección de firmas, para no hablar de que la misma firma podría ser usada para tomar represalias, como efectivamente lo fue.

Por los factores antes expuestos es esperable que en muchos centros con un bajo número de firmas, el número de votos por el SÍ pudiera ser varias veces mayor que el número de firmas. Esto se observa en los manuales pero no en los automatizados.

Si las dificultades para firmar no hubiesen existido, el comportamiento que efectivamente se observa en los centros automatizados habría sido el que cabría esperar. Firmar habría sido más o menos equivalente que votar. Efectivamente los gráficos de los votos opositores de las elecciones de 2000 contra los de 1998 que se muestran en la Fig. 9 son bastante semejantes a los de la Fig. 1

El comportamiento extraño de los automatizados, que no muestran las dificultades que hubo para firmar, se pone en evidencia si comparamos el Referendo con el resultado de otras elecciones en las que el efecto de las trabas para firmar sí es evidente. La primera parte de la fórmula que define  $k$  permite hacer esa comparación con elecciones diferentes. En las figuras Fig. 4 y Fig. 5 se muestran, para los centros que fueron automatizados en el RR, los gráficos de  $k$  contra  $f$  correspondientes a las elecciones de 1998, de 2000, a las regionales de 2004 y a un exit poll del Revocatorio. Naturalmente para cada caso se nota el cambio en el porcentaje de votos opositores, que hace que, por ejemplo, para 1998 la nube de puntos suba a medida que  $f$  baja, mientras que para las regionales de 2004 parece abrirse en abanico. Eso no es lo importante; lo que queremos destacar es que para todos los resultados, a semejanza de los centros manuales, hay una nube de puntos con  $k$  por encima de 2, o sea en los que el porcentaje de votos opositores fue varias veces superior que el porcentaje de firmas.

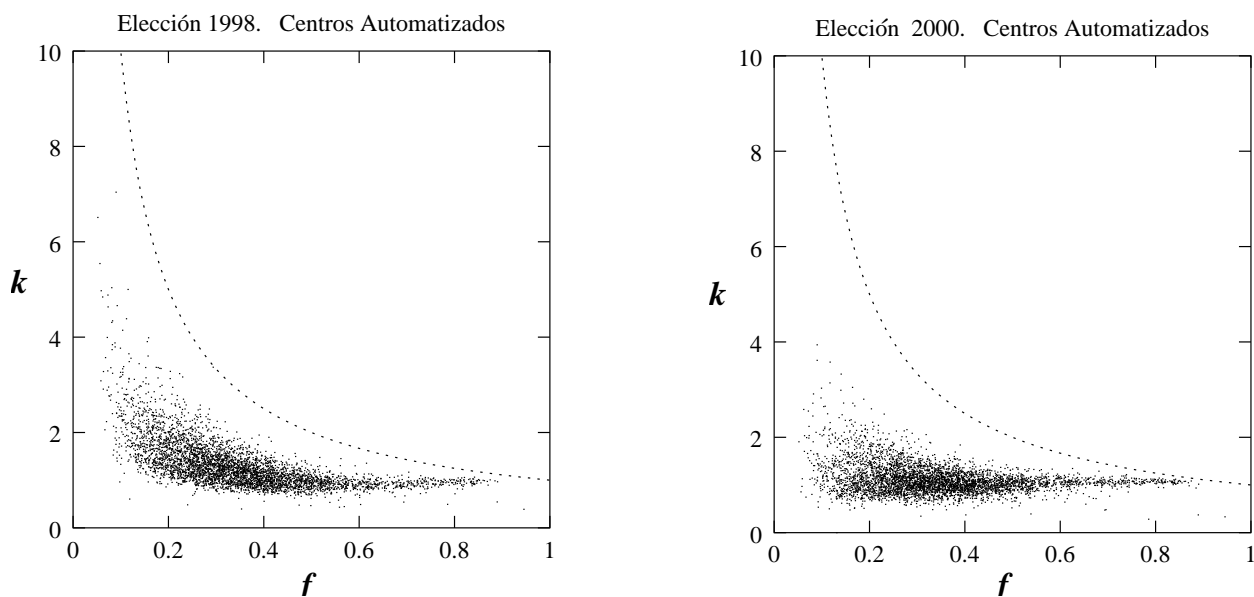


Fig. 4  $k$  vs  $f$  para las elecciones de 1998 y 2000

Es de hacer notar que las presidenciales de 2000 corresponden al momento de mayor popularidad de Chávez antes del revocatorio, no se comprende como hay tantos centros que estaban en la nube de puntos por encima de  $k = 2$  en los que la oposición bajó. Análogamente, no se comprende como la oposición subió en tantos centros en la elección regional de 2004, después de la desmoralización producida por el resultado del referendo y la campaña abstencionista. Tampoco se entiende como el exit poll, que se supone que es un muestreo del revocatorio, dé resultados sistemáticamente por encima de los oficiales.

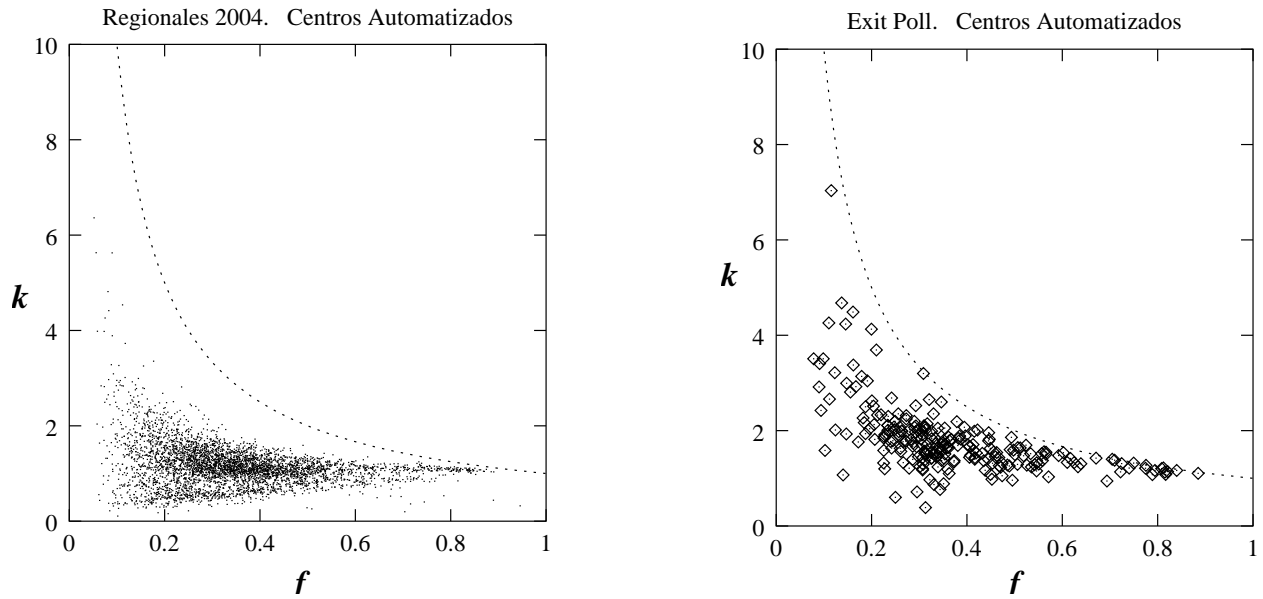


Fig. 5  $k$  vs  $f$  para las regionales de 2004 y un exit poll del RR

En el apéndice C se presenta una serie de casos de municipios en los que hubo centros automatizados y manuales.

En su trabajo Delfino y Salas justifican su hipótesis con una serie de cálculos que no es el caso repetir aquí. En cambio en la Fig. 6 vemos un histograma del número de electores en función de  $f$ . Nótese que la gran mayoría de los votos se encuentran en centros con  $f \approx 0.30$ . Esto hace que sea muy productivo un fraude como el indicado por ellos. El fraude sería más efectivo en los centros con  $f$  bajo que en los de  $f$  alto en los cuales probablemente simplemente no se efectuaría.

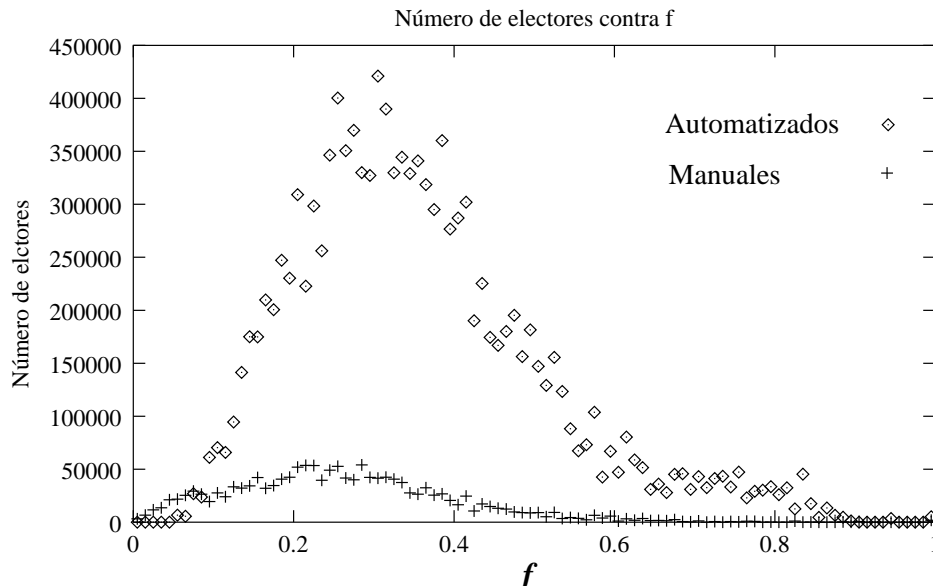


Fig. 6 Histogramas para los centros manuales y automatizados del número de electores contra  $f$

## 2. Correlación inexplicable en la auditoría en caliente

En el Referendo Revocatorio estaba previsto que se auditaría el 1% de las máquinas al terminar la elección (auditoría en caliente). Para eso se escogieron 196 máquinas (en 195 centros). Cuando llegó el

momento el mismo CNE frustró la mayoría de las auditorías pretextando razones técnicas. Al final se auditaron solamente 27 máquinas localizadas en 26 centros. Vamos a *demostrar* que las razones que frustraron las auditorías estaban relacionadas con el número de firmantes en el centro. Para esto clasifiquemos los 195 centros escogidos según la fracción de firmantes  $f$ . En la Fig. 7 vemos que los valores de  $f$  están bien distribuidos en el rango. En cambio los centros realmente auditados están principalmente en valores altos de  $f$ . ¿Qué probabilidad hay de que esto sea casual?. El promedio de los 195 valores de  $f$  es

$$f_p = 0.374345$$

y la desviación típica de esos valores es

$$\sigma_0 = 0.170167.$$

La desviación típica del promedio de 26 valores tomados al azar es

$$\sigma = \frac{1}{\sqrt{26}}\sigma_0,$$

pero cómo los valores no son totalmente independientes porque no se pueden repetir hay que corregir  $\sigma$ . Tenemos

$$\sigma = \frac{1}{\sqrt{26}}\sigma_0\sqrt{\frac{195-26}{195}} = 0.0310682.$$

El promedio de los 26 centros auditados es  $f_a = 0.542589$ , y la desviación respecto del promedio de todos los centros es

$$t = \frac{f_a - f_p}{\sigma} = 5.415316$$

Usando la distribución de Gauss se puede calcular con este valor de  $t$  que la probabilidad de que por azar el promedio sea el resultado de los centros auditados o mayor es igual a

$$\text{Probabilidad}(f_a \geq 0.54259) = 3.059 \times 10^{-8}.$$

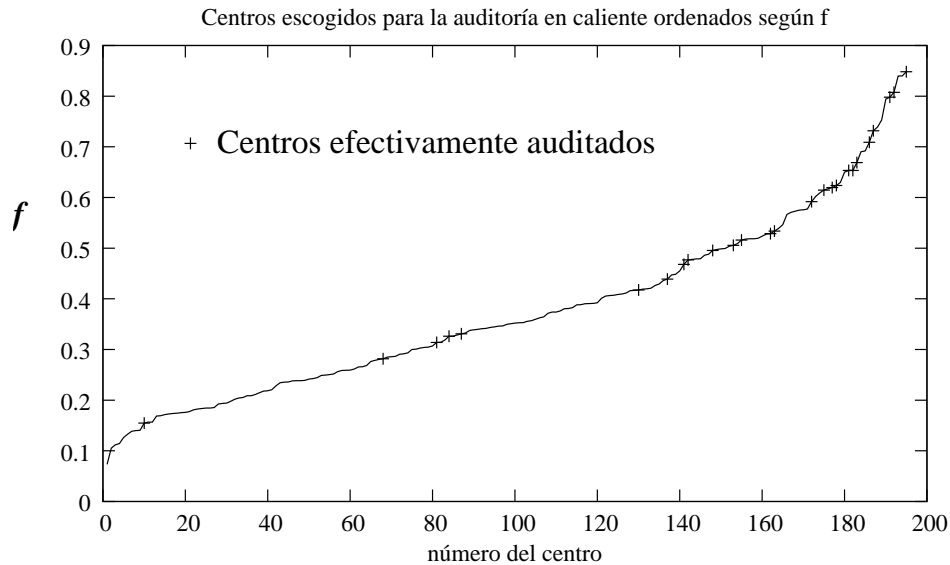


Fig. 7 Centros escogidos para ser auditados ordenados según  $f$

En la Fig. 8 se muestra un histograma de 100 columnas con el resultado de hacer 10 millones de promedios de 26 centros sacados al azar. Como se puede verificar coincide con la teoría aquí expuesta.

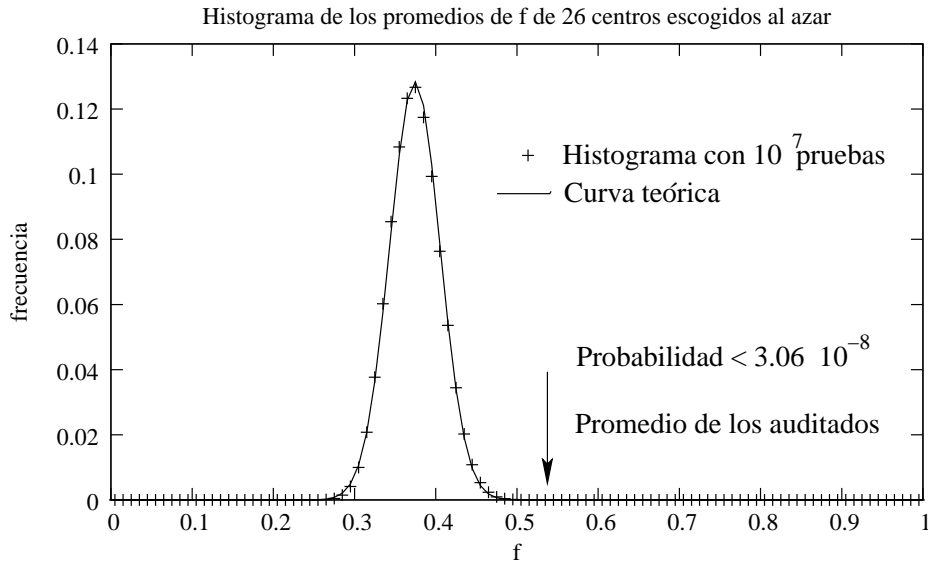


Fig. 8

En conclusión queda demostrado que **la razón por la que se auditaron unos centros y no otros está relacionada con el número de firmas en el centro.**

Ninguna razón inocente puede explicar esa correlación. ¿Qué razón técnica u organizativa puede tener que ver con el resultado del referendo o con el número de personas que firmó en un centro? Por otra parte si la hipótesis de Delfino y Salas fuese cierta habría mayor probabilidad de que en los centros con un alto valor de  $f$  no se haya efectuado el fraude, y esos son los centros que conviene auditar.

### 3. El número de votos SÍ fue inventado a partir de las firmas

Si la hipótesis de Delfino y Salas fuese correcta los resultados oficiales del Referendo Revocatorio resumidos en la Fig. 1 deberían ser falsos. ¿Habría algo en los mismos resultados que permita verificar si son verdaderos o falsos?. Después de todo a primera vista la Fig. 1 no parece muy diferente de la Fig. 9 donde se presenta el gráfico de los votos opositores de la elección presidencial de 2000 contra los de la elección de 1998. Afortunadamente la respuesta es **sí**.

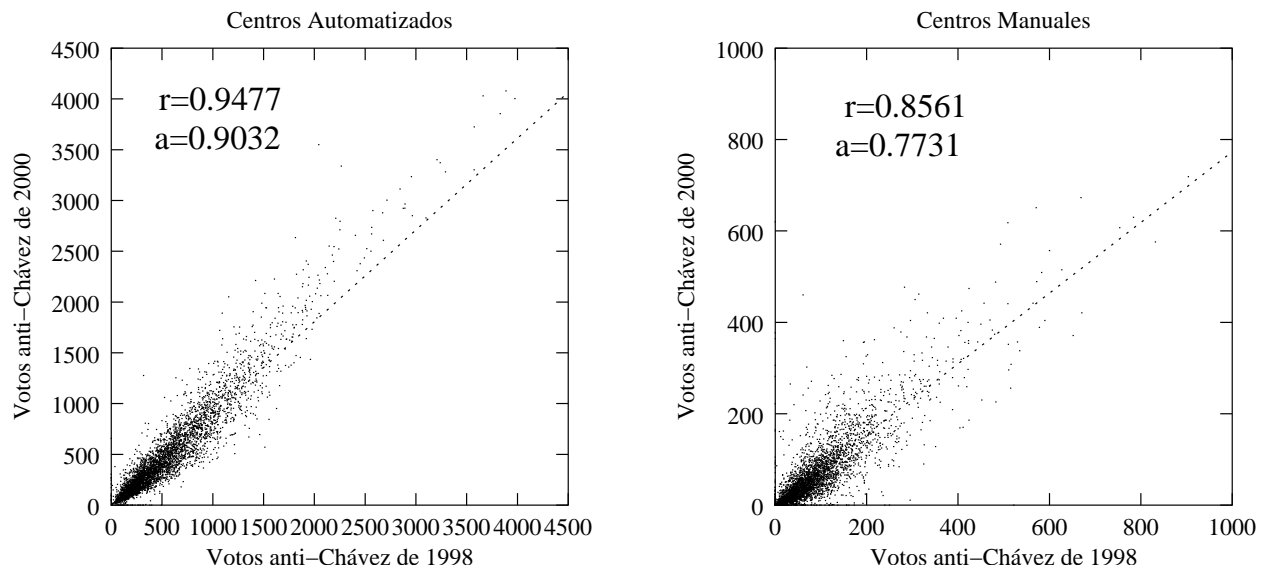


Fig. 9 Resultados de 2000 contra los de 1998

En efecto los datos de los gráficos de las figuras 1 y 9 parecen la misma cosa pero en realidad son muy diferentes. La diferencia se puede poner de manifiesto usando el test de la asimetría de las desviaciones que se explica en detalle en la sección siguiente. Ese test permite diferenciar conjuntos de pares de datos  $(x, y)$  en donde ambas variables  $x$  y  $y$  son aleatorias y fluctúan, de conjuntos en donde sólo la variable  $y$  es aleatoria y fluctúa. En resumen el test consiste en lo siguiente:

1. Se construyen sendos histogramas de las desviaciones verticales y transversales.
2. Si **ambas** variables  $x$  y  $y$  son aleatorias las desviaciones las *transversales* son *simétricas* y las *verticales* tienen *asimetría positiva*.
3. Si **sólo** la variable  $y$  es aleatoria, las desviaciones *transversales* tienen *asimetría negativa* y las *verticales* son *simétricas*.

En la Fig. 10 se ve que los votos por el SÍ de los centros manuales contra el número de firmas, efectivamente se comportan como dos variables aleatorias: desviaciones transversales simétricas y verticales asimétricas. De igual forma se comportan los datos de la Fig. 9 como se muestra en la sección 4. Por el contrario en la Fig. 11 se muestra que para los centros automatizados las desviaciones transversales tienen asimetría negativa y las verticales son simétricas. Es el comportamiento característico de una sola variable aleatoria.

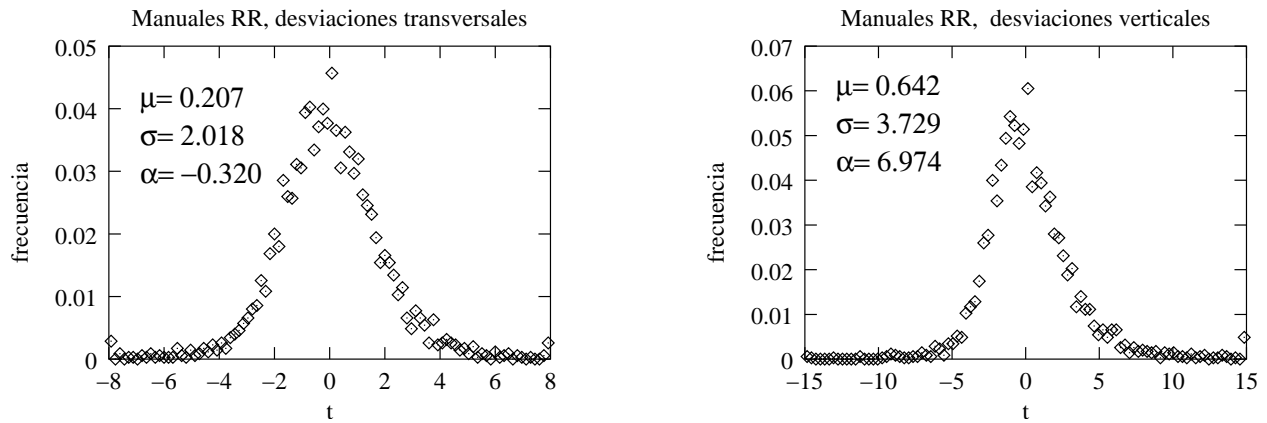


Fig. 10 Centros manuales. Desviaciones del gráfico de los votos SÍ del RR *vs* firmas

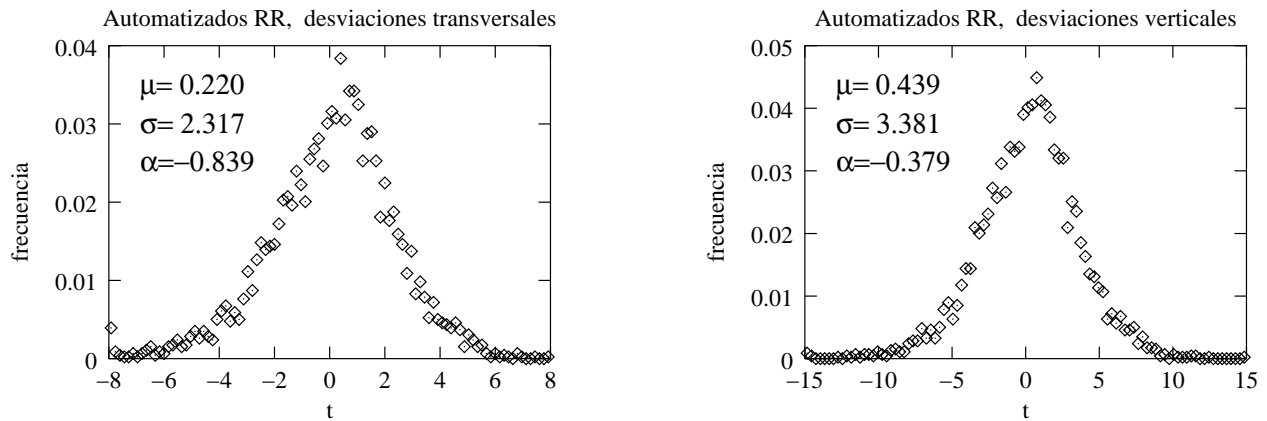


Fig. 11 Centros automatizados. Desviaciones del gráfico de los votos SÍ del RR *vs* firmas

Las cantidades que aparecen en los gráficos son el promedio  $\mu$  la desviación típica  $\sigma$  y la asimetría de Fisher  $\alpha$ . Este último parámetro mide la inclinación de la curva. Para las desviaciones transversales de los centros automatizados la asimetría de Fisher es  $\alpha = -0.839 \pm 0.105$ . Con el error estimado la probabilidad de que la asimetría de las desviaciones transversales sea mayor o igual a cero es de  $6.75 \times 10^{-16}$ .



No nos queda que concluir que para los centros automatizados no hay *dos* variables aleatorias sino *una sola*, o sea que los votos por el SÍ fueron generados usando el número de firmas, o lo que es equivalente que la hipótesis de Delfino y Salas es correcta.

La fórmula que se usó para el cálculo de los votos del SÍ pudiera haber sido algo como lo siguiente. Si llamamos  $S$  al número de votos por el SÍ y  $F$  el número de firmas, entonces

$$S = aF + tb\sqrt{aF}.$$

La variable  $t$  representa un número con distribución normal de media cero y desviación típica 1, que se obtiene con un generador de números pseudoaleatorios. La constante  $b$  se pone para que las desviaciones no sean excesivamente pequeñas y  $a$  es la constante de proporcionalidad. De los ajustes se puede estimar que  $a \approx 1.16$  y  $b \approx 3.4$ . La cosa pudiera ser más complicada, por ejemplo en vez de una única constante nacional  $a$  pudiera haber valores de  $a$  diferentes para cada región, etc.

Cuando el verdadero número de votos SI es cercano al número de firmas la fórmula podría dar un valor mayor que el verdadero. Por eso ha debido haber una condición adicional:

si  $S > c \times \text{Votos-SI-verdaderos}$  se deja el resultado verdadero.

El número  $c$  debe ser un número cercano a 1 (p. e. 0.95). De esta forma se evita hacer un fraude cuando no se necesita. La existencia de una condición de este tipo podría explicar la pequeña asimetría negativa que se observa en las desviaciones verticales. Usando una simulación se pudo estimar que el número de centros en los que el fraude fue inefectivo puede llegar al 30%, dependiendo de cual fue el verdadero porcentaje del SÍ. Nótese que estos centros están distribuidos para todos los valores del porcentaje de SÍ. Es posible entonces escoger una muestra que reproduzca el resultado falso de la elección. Naturalmente esta escogencia no sería al azar sino con un algoritmo adecuado.

#### 4. El test de la asimetría de las desviaciones transversales y verticales

Explicamos en esta sección el test de la asimetría de las desviaciones de conjuntos de datos que dos variables que tengan una relación de proporcionalidad.

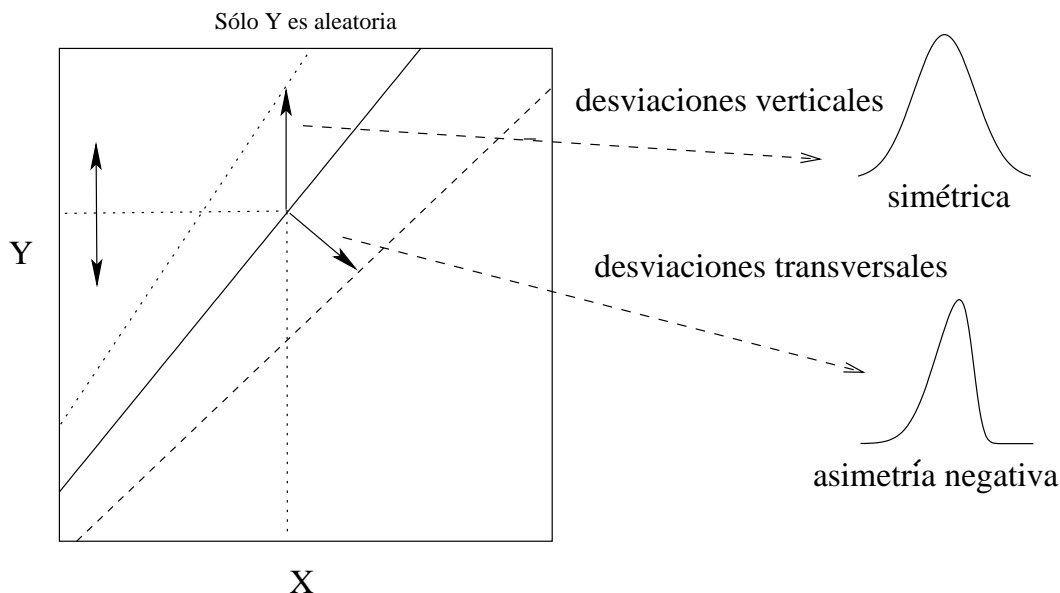


Fig. 12 Desviaciones cuando sólo  $Y$  es aleatoria

Supongamos primero que tenemos un gráfico de una variable aleatoria  $y$  en función de otra  $x$  que no lo es. Para cada valor de  $x$  el valor de  $y$  puede fluctuar alrededor de su valor esperado  $y_0 = ax$ . Las

probabilidades de los valores que  $y$  puede tomar se pueden describir típicamente con una curva en forma de campana simétrica. Para casos como los que nos interesan el ancho de la curva es proporcional a la raíz cuadrada del número. Si consideramos las desviaciones *verticales* de todos los puntos del gráfico y los acumulamos en un histograma obtendremos una curva *simétrica*. En cambio si consideramos las desviaciones *transversales* a la línea de regresión obtendremos una curva más estrecha y *asimétrica* porque las desviaciones de puntos por encima de la línea corresponden en realidad a puntos con valores esperados más pequeños que los de puntos por debajo. Esto está explicado en la Fig. 12.

En cambio si tenemos dos variables aleatorias con desviaciones típicas comparables se puede comprobar que los efectos de las dos se combinan y en general son las desviaciones *transversales* las que son *simétricas* mientras que las *verticales* son *asimétricas* pero con la inclinación contraria. Esto lo podemos ver en la Fig. 13.

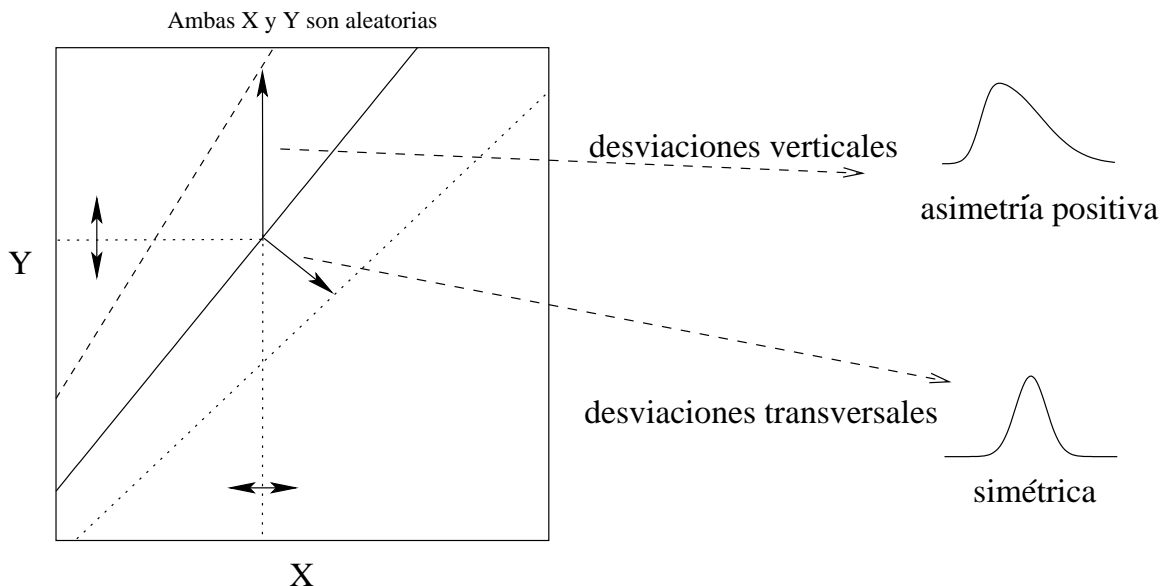


Fig. 13 Desviaciones cuando ambas  $X$  y  $Y$  son aleatorias

El razonamiento anterior es válido si las desviaciones intrínsecas de las variables son simétricas. En el caso de que las desviaciones de cada variable sean asimétricas esto se reflejará en las asimetrías de las desviaciones verticales y transversales. Por ejemplo en el caso de una sola variable aleatoria, si esta tiene una asimetría positiva esto aparecerá en las desviaciones verticales, y las desviaciones transversales tendrán una asimetría menos positiva, lo cual podría confundirse con el comportamiento de dos variables aleatorias. Sin embargo si la única variable aleatoria tiene una asimetría negativa, las desviaciones transversales tendrán una asimetría aún más negativa; en este caso no hay confusión posible. Nótese que en el caso de las dos variables aleatorias las desviaciones transversales tienden siempre a ser más simétricas que las verticales, por consiguiente nunca darán el comportamiento típico de una sola variable aleatoria.

En las figuras 14 y 15 podemos ver los histogramas de simulaciones con una y dos variables aleatorias; las simulaciones se comportan del modo previsto. Las desviaciones normalizadas  $t$  de los puntos  $(x, y)$  se calculan con la fórmula siguiente

$$t = \frac{y - ax}{s}$$

donde  $a$  es la constante de proporcionalidad y  $s$  es

$$s = \sqrt{ax}$$

para las desviaciones verticales y

$$s = \sqrt{a^2x + y}$$

para las desviaciones transversales.

La constante  $a$  se obtuvo con un cálculo de mínimos cuadrados usando como peso

$$w = \frac{x^2 + y^2}{(x + y)xy}.$$

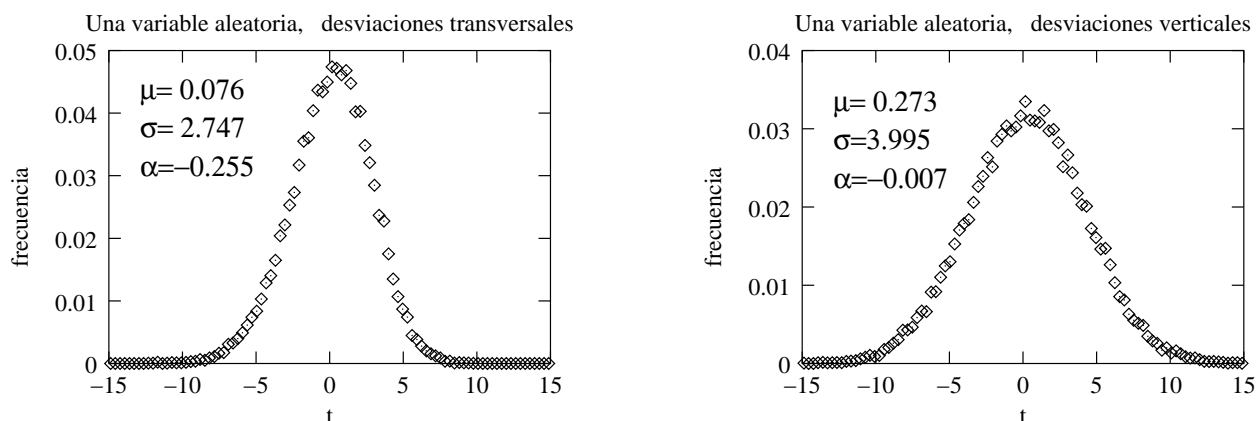


Fig. 14 Simulación con 40000 puntos. Sólo Y es aleatoria

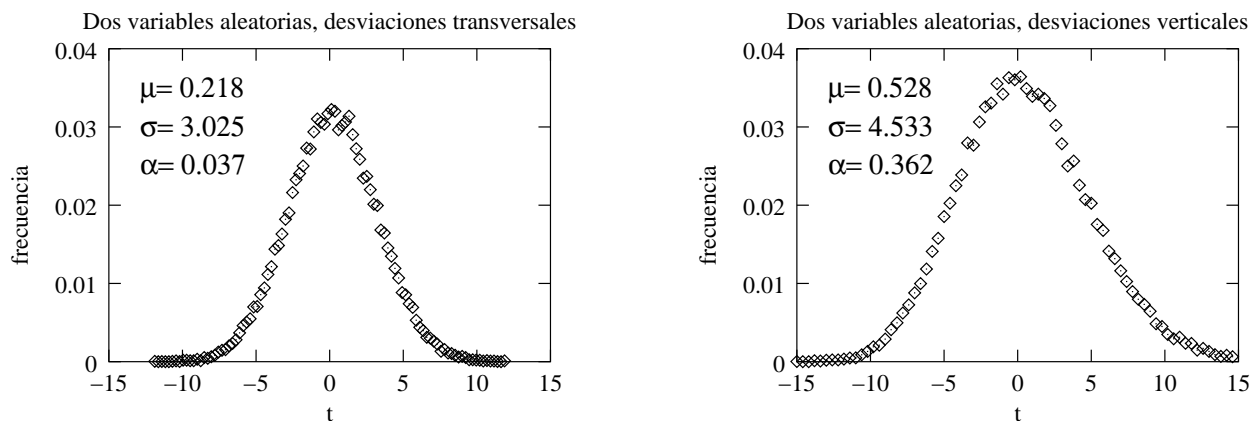


Fig. 15 Simulación con 40000 puntos. Ambas X y Y son aleatorias

En la Fig. 16 se grafican las desviaciones del número total de votos contra el número de electores inscritos en el REP para los centros automatizados del RR. Este es un caso de una sola variable en el que se presenta una asimetría intrínseca negativa. La asimetría es probablemente debida a una distribución geográfica no-uniforme de la abstención. Las desviaciones verticales tienen asimetría negativa y las transversales la tienen aún mayor. Otro caso de una sola variable es el del número de firmas contra el número de electores inscritos, también para los centros automatizados del revocatorio. Las desviaciones que se presentan en la Fig. 17 son intrínsecamente positivas. La asimetría es debida a la existencia de una población de opositores “duros” que firmó a como diera lugar. La existencia de estos electores aparece como la pequeña elevación para  $f$  alto en la Fig. 6. Las desviaciones transversales son ligeramente positivas y las verticales bastante más, por lo que este caso podría confundirse con un caso de dos variables. Por el contrario en la Fig. 18 se muestra el número total de votos contra las firmas de los centros automatizados del RR. A pesar de las asimetrías intrínsecas contrarias, en este caso el comportamiento es indudablemente el correspondiente a dos variables aleatorias: desviaciones transversales ligeramente negativas y verticales positivas. Los gráficos correspondientes a estas desviaciones se pueden ver en el apéndice B.

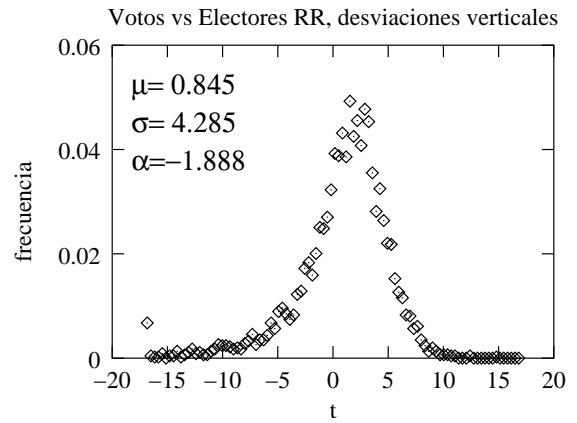
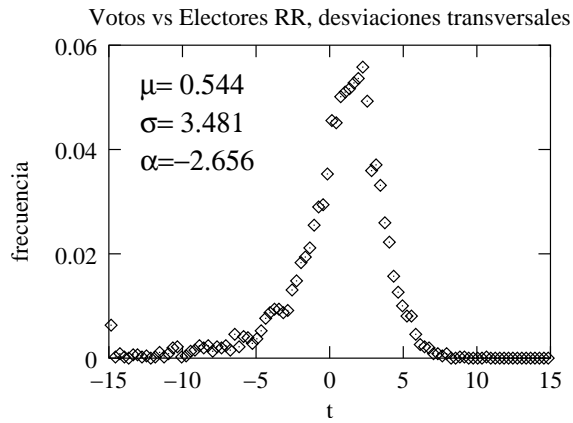


Fig. 16 Centros automatizados. Desviaciones del gráfico de votos totales *vs* número de electores del RR

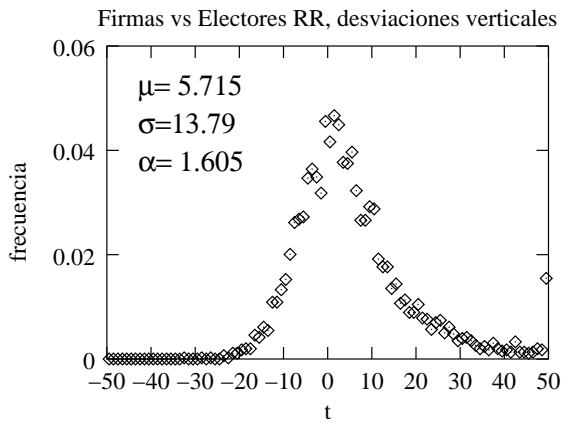
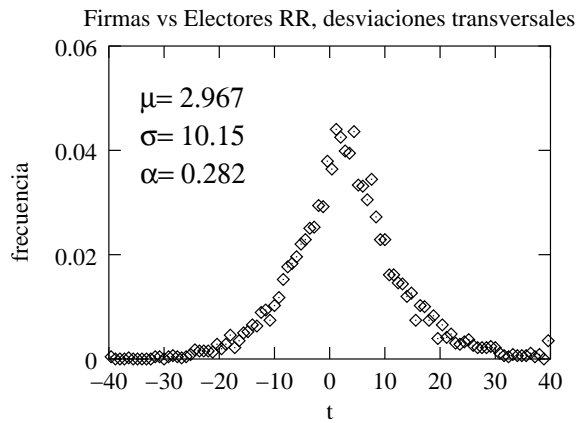


Fig. 17 Centros automatizados. Desviaciones del gráfico de firmas *vs* número de electores del RR

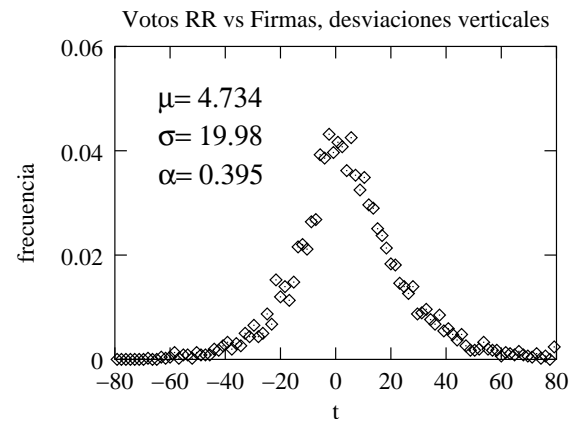
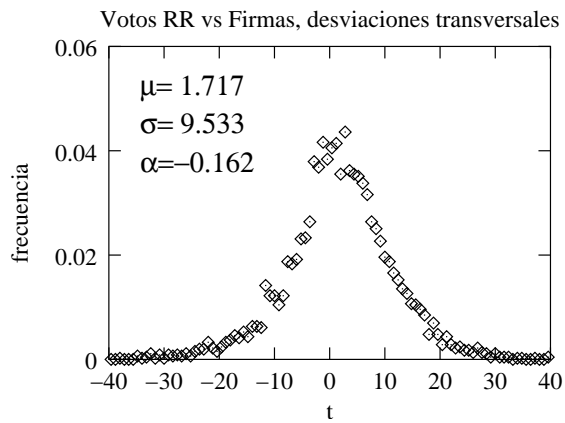


Fig. 18 Centros automatizados. Desviaciones del gráfico de votos totales del RR *vs* número de firmas

En las figuras 19 y 20 podemos ver el histograma de las desviaciones de los votos opuestos a Chávez de la elección de 2000 contra los de 1998. Se comportan como se espera para dos variables aleatorias: desviaciones transversales simétricas y verticales con asimetría positiva. Lo mismo sucede con los resultados de los centros manuales del Referendo Revocatorio (Fig. 10).

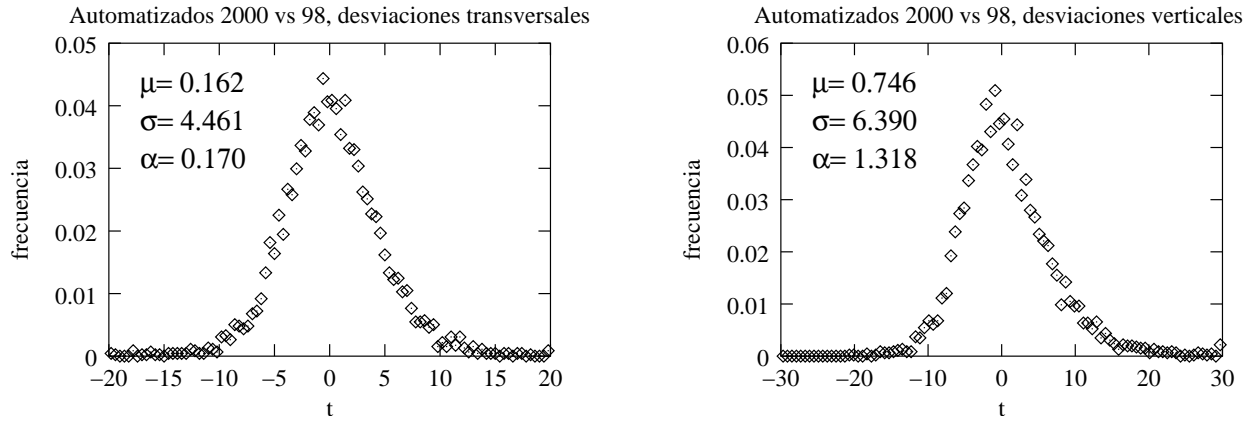


Fig. 19 Centros automatizados. Desviaciones del gráfico de votos anti-Chávez de 2000 vs votos anti-Cháves de 1998

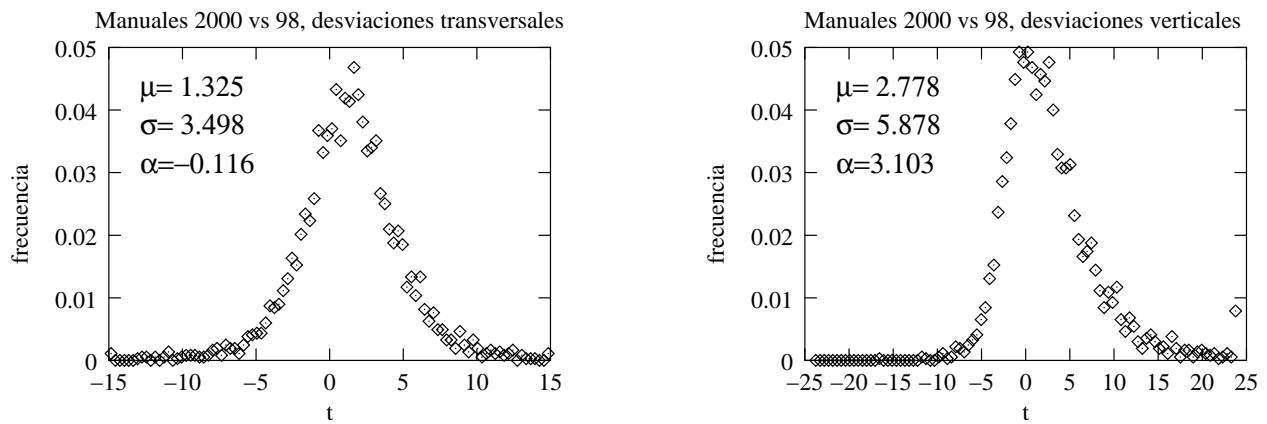


Fig. 20 Centros manuales. Desviaciones del gráfico de votos anti-Chávez de 2000 vs votos anti-Cháves de 1998

|                               | $\alpha$ transversales | $\alpha$ verticales | # variables |
|-------------------------------|------------------------|---------------------|-------------|
| Votos totales vs electores RR | -2.656                 | -1.888              | 1           |
| Firmas vs electores RR        | 0.282                  | 1.605               | 1           |
| Votos totales RR vs Firmas    | -0.162                 | 0.395               | 2           |
| Manuales 2000 vs 98           | -0.116                 | 3.103               | 2           |
| Automatizados 2000 vs 98      | 0.170                  | 1.318               | 2           |
| Manuales RR vs firmas         | -0.320                 | 6.974               | 2           |
| Automatizados RR vs firmas    | $-0.839 \pm 0.105$     | $-0.379 \pm 0.058$  | 1           |

Tabla I. Asimetrías de Fisher

En Fig. 11 se presentan los resultados de los centros automatizados del RR. Al contrario de los casos anteriores el histograma de las desviaciones transversales tiene *asimetría negativa*, y el de las desviaciones verticales es simétrico, el comportamiento característico de *una sola variable aleatoria*. En la tabla I se presentan los valores de la asimetría de Fisher  $\alpha$ . Los errores de los valores para el caso de los centros automatizados del RR son desviaciones típicas que se estimaron con un cálculo de Monte Carlo como se explica en el apéndice A. No hay duda de que la asimetría de las desviaciones verticales no es positiva y que la de las transversales es negativa.

## 5. Comportamiento extraño de las encuestas a salida de urna.

En el día del referendo se efectuó una encuesta a salida de urna (exit polls) asesorada por Penn, Schoen and Berland Associates. Esta encuesta dio ganador al SÍ por amplio margen. Hay tres tipos de error que puede tener esta clase de encuesta:

- 1- El conjunto de centros electorales escogidos para la encuesta no representa fielmente al universo de electores. Este tipo de error afecta el resultado global de la encuesta, pero no los resultados en cada centro.
- 2- Error sistemático de buena fe. Por alguna razón en el proceso mismo de la encuesta no se muestrean de igual forma los diferentes tipos de electores por diferencias en su comportamiento. En una situación de mucha polarización político-social como la venezolana esto no sería raro. Por ejemplo puede haber diferencias en la propensión a dejarse entrevistar o en las horas en las que se va a votar.
- 3- Manipulación. Pueden estar manipulados los resultados oficiales de la elección o también los resultados de la encuesta. La encuesta puede ser manipulada por los encuestadores o por los mismos entrevistados en caso de que en un alto porcentaje decidan responder falsamente.

Cualquiera que sea la razón los errores del segundo tipo producen una diferencia sistemática entre la predicción de la encuesta y el resultado verdadero, diferencia que se manifiesta en todos los centros. Llamemos  $x$  la fracción de votos SÍ y  $q$  la razón entre las probabilidades de encuestar los electores del NO y los del SÍ. Si  $q = 1$  hay igual probabilidad de encuestar unos y otros, si  $q = 0.5$  hay el doble de probabilidad de encuestar a un opositor que a un partidario del NO. Con un cálculo simple se puede demostrar que la diferencia  $\Delta$  entre la fracción de votos SÍ predicha por la encuesta y el resultado verdadero es, en promedio

$$\Delta = \frac{(1-x)x(1-q)}{x+q(1-x)}.$$

Nótese que para los extremos  $x = 0$  y  $x = 1$  el error se anula. Si todos votan SÍ o todos votan NO la encuesta no se puede equivocar.

En la Fig. 21 se muestran en función de la fracción de votos SÍ las diferencias entre las predicciones de la encuesta y los resultados oficiales en los 239 centros automatizados que fueron encuestados. Lo primero que hay que notar es lo grande que son el error sistemático y la dispersión de los datos. Hay un solo punto por debajo de  $\Delta = -0.1$  mientras que la mayoría está por encima de  $\Delta = 0.1$ . A simple vista se nota que mientras para  $x \approx 1$  efectivamente las diferencias son pequeñas, no sucede lo mismo en el otro extremo del gráfico, más bien las diferencias parecen crecer. Esto se evidencia en el coeficiente de correlación negativo ( $r = -0.3258$ ) y en el hecho de que un ajuste por mínimos cuadrados de la fórmula prevista ( $q = 0.452$ ) tiene una suma normalizada de cuadrados igual a  $\chi^2 = 0.0154$  que es bastante mayor que la de una regresión lineal  $y = ax + b$  que tiene  $\chi^2 = 0.0139$ ,  $a = -0.211 \pm 0.040$ ,  $b = 0.260 \pm 0.018$ . Sin ninguna duda *los errores de las encuestas no siguen la fórmula prevista para los errores sistemáticos de buena fe*. Podemos inferir que hubo manipulación de la encuesta o del referendo.

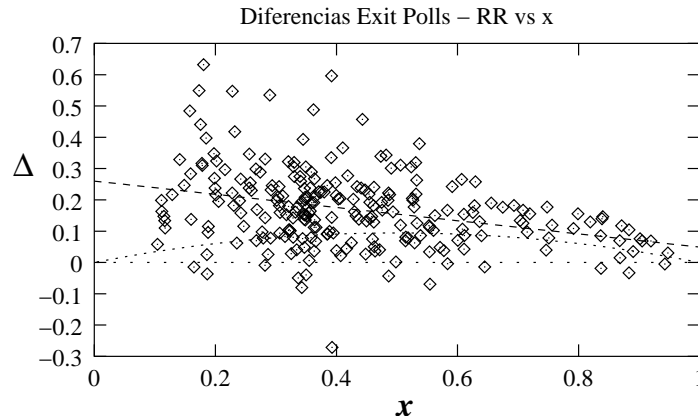


Fig. 21 Diferencias entre las fracciones de votos SÍ estimadas por las encuestas y los resultados oficiales *vs* la fracción  $x$  de votos SÍ. La línea curva es la curva teórica de los errores sistemáticos que mejor se ajusta a los datos ( $q = 0.452$ ). La línea recta es la regresión lineal cuyos parámetros se dan en el texto.

En la Fig. 22 se muestran las diferencias entre las encuestas y los resultados oficiales en función de la fracción de firmas  $f$ . En este caso el error de las encuesta no tiene que ser pequeño para  $f \approx 0$ , porque un número pequeño de firmas no implica necesariamente un número pequeño de votos por el SÍ. En cambio

un número alto de firmas sí implica un número alto de votos SÍ por lo que es de esperarse que los errores sean pequeños, cosa que efectivamente se observa. Lo que no tiene ninguna explicación es que los errores de la encuesta tenga una correlación negativa con la fracción de firmas. La regresión lineal  $y = ax + b$  da  $a = -0.244 \pm 0.045$  y  $b = 0.260 \pm 0.018$ . Nótese que la línea pasa muy cerca de cero para  $f = 1$ . La probabilidad de que no haya correlación ( $a = 0$ ) es de  $2.94 \times 10^{-8}$ . Esto es una confirmación adicional de que hubo una manipulación, porque no es imaginable un error sistemático de buena fe, o sea una diferencia en la probabilidad de entrevistar a unos y otros, que dependa del número de firmas que se recogieron en el centro.

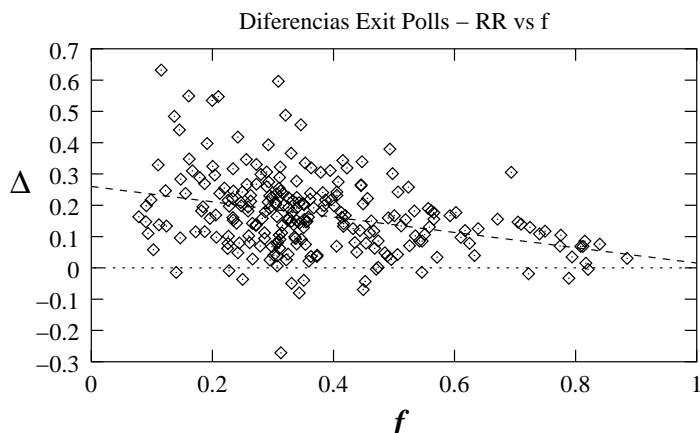


Fig. 22 Diferencias entre las fracciones de votos SÍ estimadas por las encuestas y los resultados oficiales *vs* la fracción de firmas  $f$ . La línea recta es la regresión lineal cuyos parámetros se dan en el texto.

Si usamos, para calcular las diferencias con las encuestas, la fracción de votos opositores en la regionales de 2004 en vez de los resultados del RR, obtenemos el gráfico que se muestra en la Fig. 23. Tres aspectos son dignos de mención:

- 1- La dispersión de los datos no es mucho mayor que la del RR; o sea que las encuestas predicen casi tan mal los resultados del RR como los de las regionales de 2004.
- 2- No hay correlación con el número de firmas. La regresión lineal da  $a = -0.072 \pm 0.063$  y  $b = 0.212 \pm 0.026$ , pero tan bueno como ese es el ajuste con una constante  $a = 0$  y  $b = 0.186 \pm 0.010$ . La probabilidad de que sea  $a \geq 0$  es del 12.65%, cosa que es perfectamente posible.
- 3- La diferencia constante del 18.6% entre los resultados de las encuestas y los de las regionales de 2004 pueden simplemente reflejar la disminución de la participación de la oposición en relación con el RR.

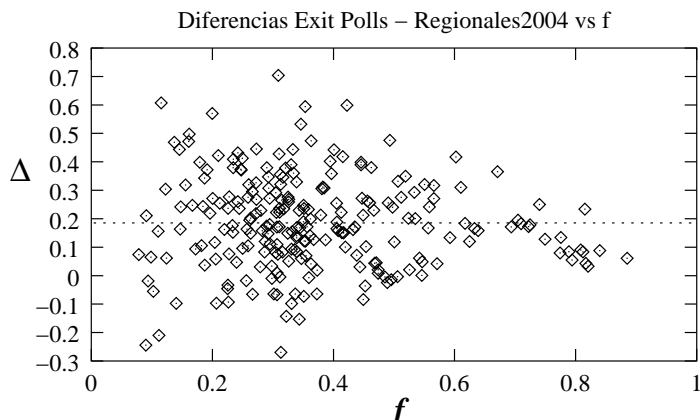


Fig. 23 Diferencias entre las fracciones de votos SÍ estimadas por las encuestas y los resultados de las regionales de 2004 *vs* la fracción de firmas  $f$ . La línea recta horizontal es la regresión con la constante.

En resumen,

- 1- El error sistemático de las encuestas no se comporta como un error sistemático de buena fe. Esto indica que hubo una manipulación de los resultados de las encuestas o bien de los resultados oficiales del RR.
- 2- El error de las encuestas está correlacionado negativamente con las firmas, lo que indica que la susodicha manipulación está relacionada con las firmas.
- 3- No son los resultados de las encuestas los que están correlacionados con las firmas, como indica la ausencia de correlación entre  $f$  y las diferencias de las encuestas con las regionales de 2004.

Conclusión:

*Los resultados oficiales del Referendo Revocatorio están manipulados de alguna manera relacionada con las firmas.*

Es un resultado consistente con la hipótesis de Delfino y Salas.

Si asumimos que en realidad no hubo tal error sistemático de las encuestas y que por lo tanto la regresión lineal del gráfico  $\Delta$  contra  $f$  mide la diferencia promedio entre el resultado oficial y el verdadero, podemos estimar cuanto es la diferencia total de votos calculando centro por centro la diferencia usando el valor de  $f$ . El resultado es de 1.370.000 votos con un error del 20%. También podemos calcular el resultado del RR usando el porcentaje de voto opositor en los centros automatizados de las regionales de 2004 (40.58%) y la diferencia de porcentaje estimada por las encuestas (18.6%). Esto da para los centros automatizados un 59.18% para el SÍ en el revocatorio. Como para esos centros El resultado oficial fue de 42.13% se obtiene una diferencia del 17.05% que corresponde a 1.457.000 votos con un error del 6%, resultado totalmente compatible con el anterior.

## 5. Conclusión

Hemos presentado en este trabajo tres evidencias independientes a favor de la hipótesis de Delfino y Salas. Demostramos que la causa que determinó en cuales de los 195 centros escogidos inicialmente se efectuó la auditoría en caliente está relacionada con el número de firmas recogidas en el centro, lo cual indica un comportamiento fraudulento del CNE. Demostramos que la encuesta a salida de urna de Penn, Schoen and Berland Associates presenta anomalías que solo pueden ser explicadas si los resultados oficiales del Referendo Revocatorio fueron manipulados de alguna manera relacionada con las firmas. Y lo más importante, aplicándole a los pares de datos (firmas,votos-SÍ) el test de la asimetría de las desviaciones verticales y transversales pudimos determinar que para los centros automatizados la variable ‘firmas’ se comporta como si no fuese aleatoria, indicando que en realidad en tales centros el número de votos SÍ fue generado con algún proceso aleatorio a partir del número de firmas en el centro. Sólo el número total de votos de las máquinas es real.

No cabe duda de que en el Referendo Revocatorio de 2004 hubo un fraude masivo en las máquinas de votación. Este fraude fue hecho por el mismo CNE o con la anuencia de ese organismo. El fraude pudo haber sido efectuado mediante la comunicación máquina-CNE o estar incluido en el software de cada máquina, por ejemplo poniendo topes preestablecidos al número de votos SÍ.

Nuestro resultado complementa los de Delfino y Salas y es compatible con los del grupo de la USB que estudió la comunicación máquinas-CNE [2].

El fraude que hemos descubierto aquí se refiere a los votos por el SÍ, pero no se excluye que haya podido haber simultáneamente en algunos centros un fraude en el NO con votantes virtuales.

## Referencias

1. G. Delfino y G. Salas (2005), “Analysis of the venezuelan Presidential Recall Referendum of 2004 and the relationship of the official results and the signatures requesting it in computerized centers”, Sometido a consideración en *Statistical Science*, Preprint en <http://www.gentederedes.org>
2. Roberto Giusti, “El 70% de las máquinas fue manipulado”, *El Universal*, 27/11/2005.

## Apéndice A. Datos y métodos

Para hacer este trabajo se usó la base de datos puesta gentilmente a disposición por Delfino y Salas.



De los 4755 centros automatizados en 177 hubo fallas en alguna de las máquinas. Estos centros se excluyeron tanto de los automatizados como de los manuales. Los resultados de estos centros son algo extraños. De 28 de ellos (el 16%) nunca se supo el resultado, Siete centros tuvieron  $k < 0.5$  (el 4%) comparado con 11 en los otros 4588 centros (el 0.24%) y en 2 el número de firmas fue mayor que el número total de votos ( $f > 1$ ). Esto indica que muchos electores se vieron impedidos de votar. Sin embargo en 2 centros (el 1.1%)  $k$  fue mayor que 2 contra 7 en los otros 4588 (el 0.15%).

Los errores de la asimetría de Fisher  $\alpha$  de las desviaciones de los votos SÍ contra firmas que aparecen en la tabla I son desviaciones típicas que fueron estimadas de la siguiente forma. Se escogieron al azar  $M$  subconjuntos de  $N$  puntos. Se calcula para cada subconjunto el valor de la asimetría  $\alpha$ . Se determina la desviación típica con los  $M$  valores de  $\alpha$ . Se hicieron cálculos para  $N = 200, 400, 800, 1600, 2300$  y  $3000$ . Los correspondientes valores de  $M$  fueron 100000, 100000, 50000, 25000, 20000 y 16000. Los valores para  $N = 4588$  se obtuvieron por extrapolación como se ve en la Fig. 24. Los valores son  $\sigma = 0.105$  para las desviaciones transversales y  $\sigma = 0.058$  para las verticales.

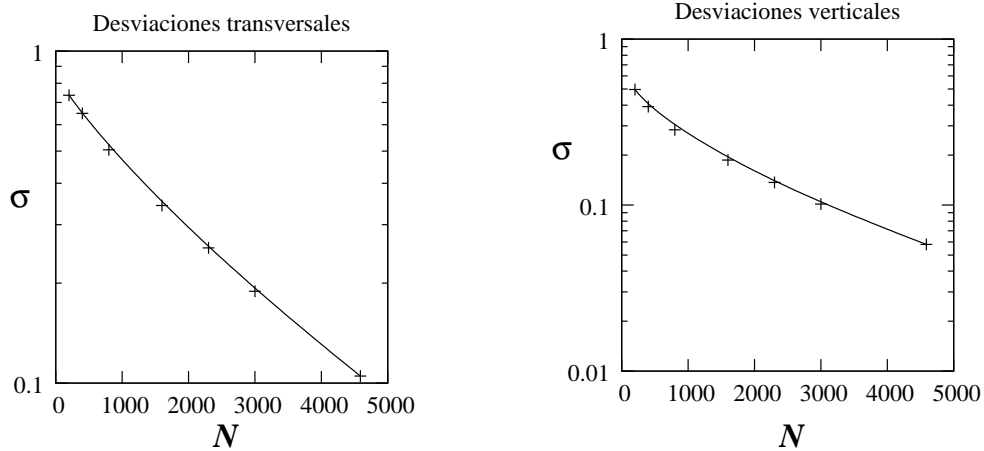


Fig. 24 Desviaciones típicas  $\sigma$  en función del tamaño  $N$  del subconjunto. El valor para  $N = 4588$  se obtiene por extrapolación.

## Apéndice B.

En este apéndice se presentan algunos de los gráficos cuyas desviaciones se usaron como ejemplos en la sección 4.

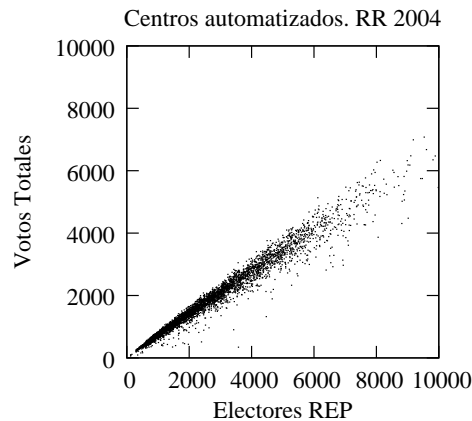


Fig. 25 Centros automatizados. Votos totales contra electores en el REP

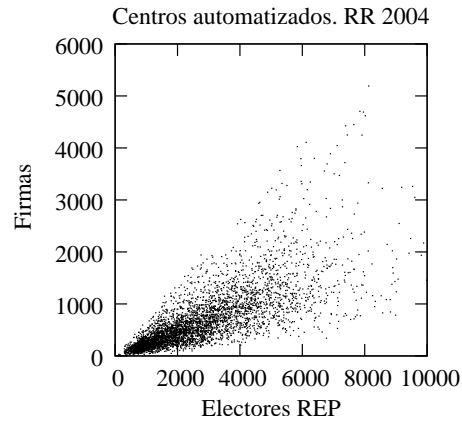


Fig. 26 Centros automatizados. Firmas contra electores en el REP

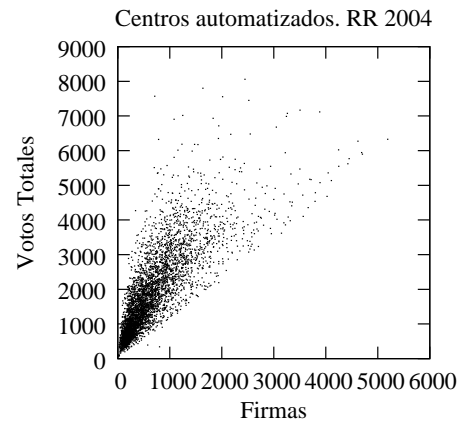


Fig. 27 Centros automatizados. Votos totales contra Firmas

### Apéndice C.

En este apéndice se presentan varios municipios en los que hay centros manuales y automatizados. En los casos del municipio Naguanagua del estado Carabobo (Fig. 28), del municipio Crespo de Lara (Fig. 29) y en la parroquia Vásquez de Maracaibo (Fig. 30) se nota que mientras en los centros automatizados la fracción de votos por el SÍ es proporcional a las firmas, en los centros manuales sigue una línea recta del tipo  $y = af + b$ , con un término constante  $b$  positivo. La constante  $b$  representa la fracción de votos SÍ que habría en un centro aún en el caso de que no hubiera firmas.

El caso de Duaca (Fig. 29) es interesante porque en ese municipio hubo cuatro centros en los que la automatización falló y que terminaron siendo escrutados manualmente. De estos centros el que tiene la menor fracción de firmas  $f = 0.1355$  tiene un  $k = 2.266$ . O sea que el número de votos SÍ más que duplicó el número de firmas. El comportamiento de este centro se corresponde con el de los centros manuales y no con el de los que fueron escrutados automáticamente en ese mismo municipio.

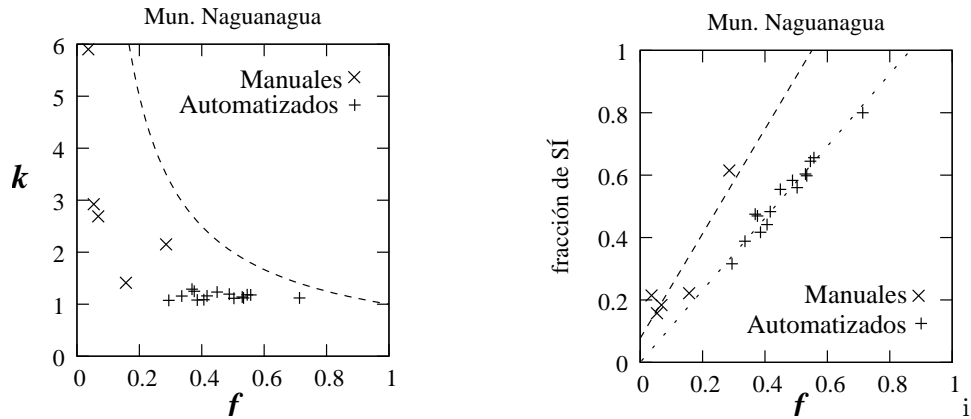


Fig. 28 Municipio Naguanagua, centros automatizados y manuales

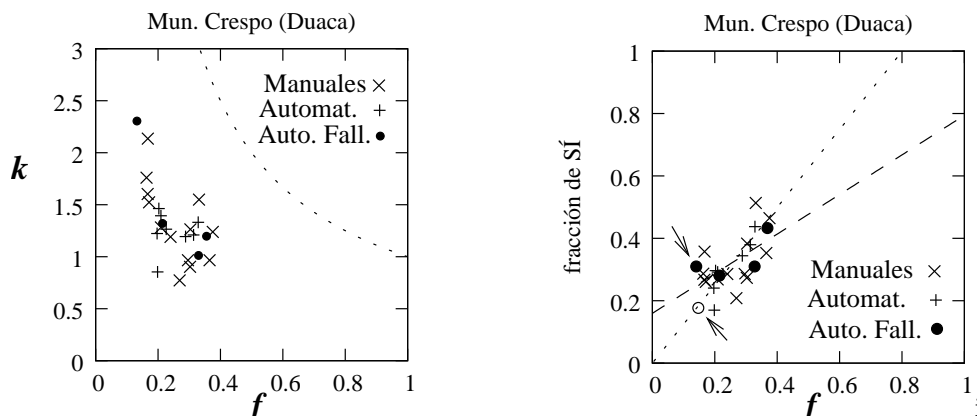


Fig. 29 Municipio Crespo, estado Lara. Centros automatizados y manuales. El punto negro indicado con la flecha corresponde al centro en le que falló la automatización con  $k = 2.27$ . El punto blanco indica la posición que hubiera debido tener el punto negro señalado de cumplirse la proporcionalidad con el número de firmas.

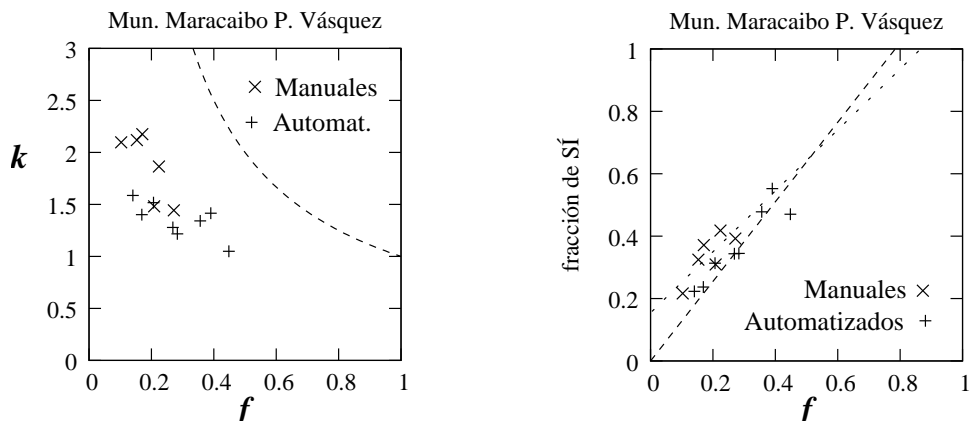


Fig. 30 Parroquia Vázquez de Maracaibo

Finalmente en la Fig. 31 presentamos el caso del municipio Rangel del estado Mérida (Mucuchíes). Dos de los 7 centros automatizados (de 4588) que tienen  $k > 2$  pertenecen a ese municipio. Como caso casi único aquí la fracción de votos SÍ no es proporcional a la fracción de firmas, sino que sigue una línea recta con termino constante positivo, comportamiento que no es raro en los centros manuales.

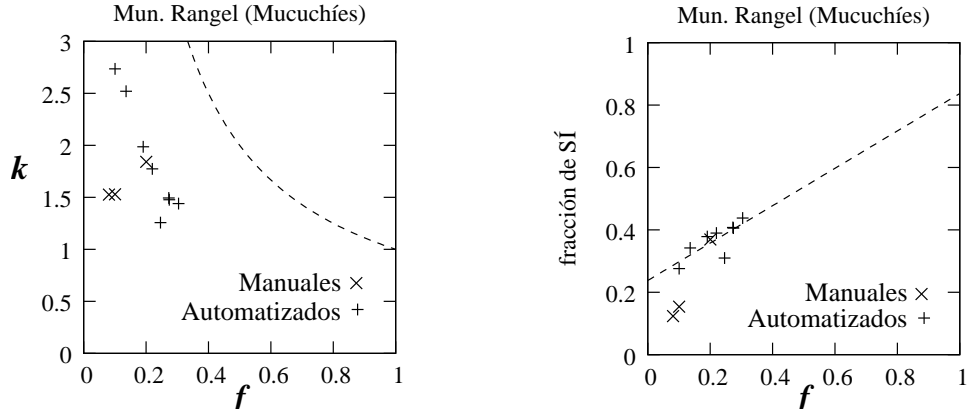


Fig. 31 Municipio Rangel (Mucuchies)

Para verificar que estos resultados no sean simplemente fluctuaciones, se hizo la siguiente estadística. Se determinaron los municipios (o parroquias en el Distrito Capital y en el estado Vargas) en los que el número de centros manuales y el de centros automatizados eran ambos mayores o iguales a 5. Son 135 los municipios que cumplen la condición. En cada municipio se calculó las regresiones lineales para los centros manuales y para los centros automatizados. Se calculó la diferencia entre los términos constantes de ambas regresiones.

$$\Delta b = b_{\text{manuales}} - b_{\text{automatizados}}$$

La tabulación de esas diferencias se presenta como histograma en la Fig. 32. Se nota claramente que hay un sesgo hacia las diferencias  $\Delta b$  positivas. Esto confirma que sistemáticamente los centros manuales tienen términos constantes mayores que los de los automatizados de su mismo municipio. Como presumiblemente los términos constantes son el reflejo de las trabas para firmar, concluimos que, inexplicablemente, los centros automatizados fueron afectados mucho menos que los manuales por dichas trabas.

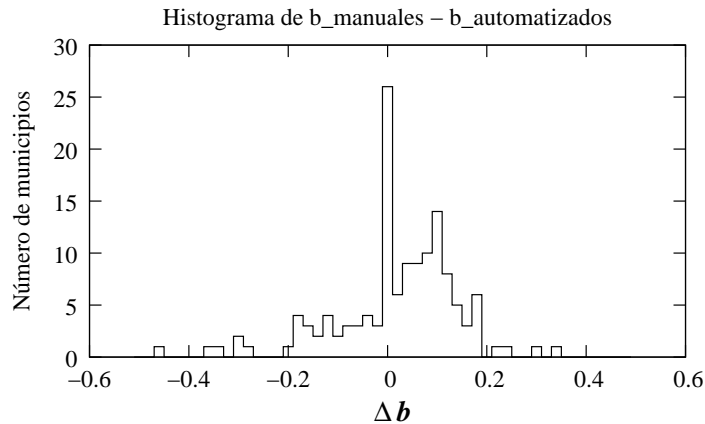


Fig. 32 Histograma de las diferencias entre los términos constantes de los centros manuales y automatizados