



Elecciones Presidenciales 2006: Análisis Estadístico de los Resultados Publicados por el PREP

Agosto 2006

Reflexiones a Tratar

- ¿Por qué los distintos conteos rápidos pudieron mostrar resultados en dos direcciones?
 - Diseño aleatorio simple
 - Diseño estratificado
 - Diseño sistemático sobre lista ordenada (Estratificación implícita)
- ¿Realizar cortes en el PREP implica analizar muestras representativas?
 - Elección de cortes “interesantes”
 - Pruebas de bondad de ajuste (Ji-cuadrada)
 - Análisis de cada uno de los cortes
- ¿Más cruces entre las preferencias indican mayor certidumbre sobre las cifras?
 - Simulación de captura inversa
- ¿La ley de Benford se puede aplicar a los votos en las casillas?
- Conclusiones
- Anexos



¿Por qué los conteos rápidos pudieron mostrar resultados en dos direcciones?



Metodología

- Con base en las cifras de las actas de las casillas que se recibieron, se procesaron y se contabilizaron en el PREP 2006, por parte del IFE, se realizaron dos ejercicios:
 - 1.Utilizando un diseño muestral aleatorio simple irrestricto, se generaron mil muestras de tres tamaños distintos:
 - n=1,000
 - n=2,000
 - n=5,000
 - 2.Utilizando un diseño muestral estratificado se generaron mil muestras
 - n=1,000

Cifras al Cierre del PREP 2006

Actas Contabilizadas:	117,287
Actas Inconsistentes:	11,184
VMRE:	300
Total Actas Procesadas:	128,771

Marco muestral

Total Nacional + VMRE: 130,788

98.45% de las Actas Totales*

Fuente: IFE

* El 1.55% restante corresponde a casillas sin actas y a actas no recibidas

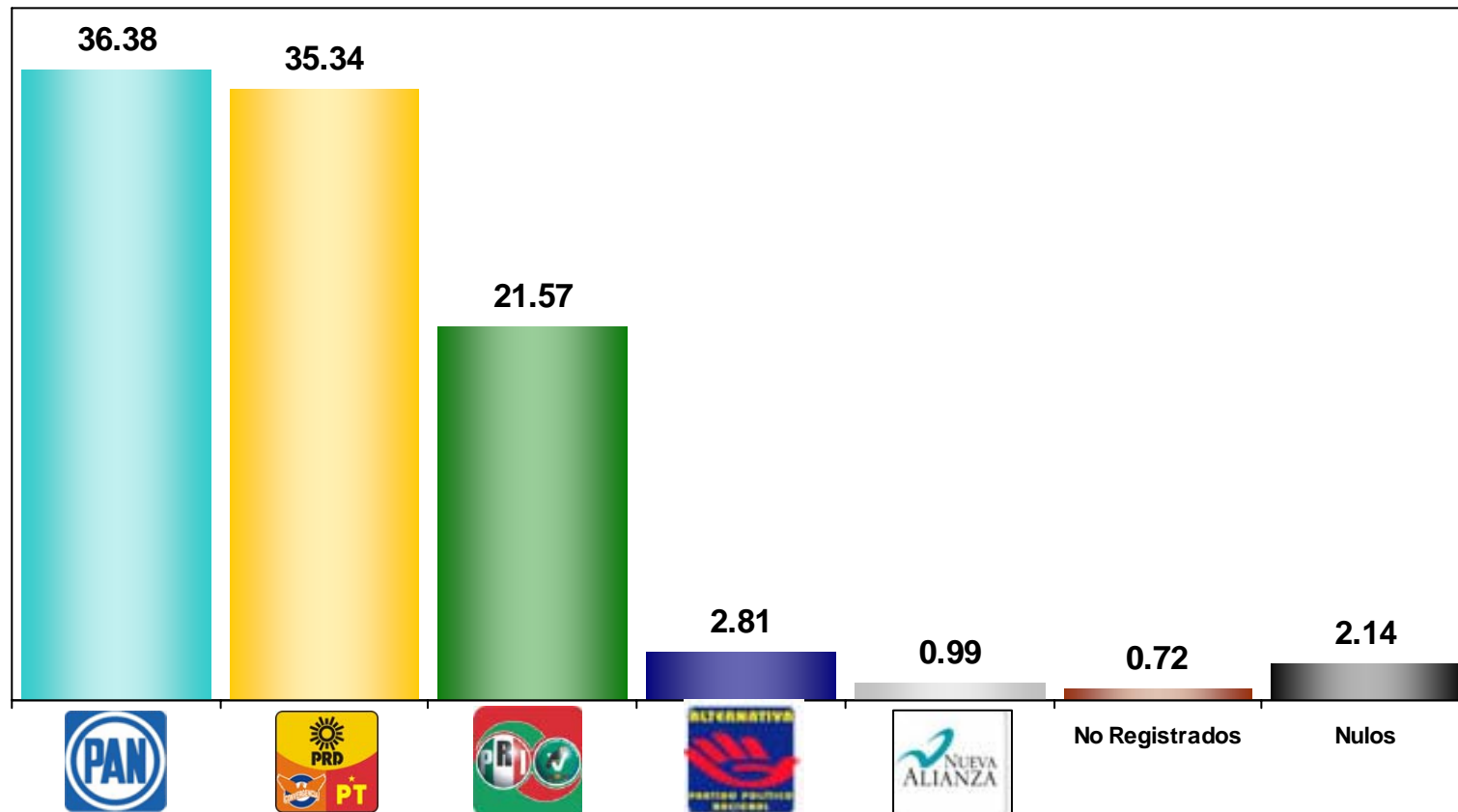


Elección 2006

Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

Al cierre del PREP, el PAN mantenía una ventaja de 1.04 puntos

- Resultados PREP 2006 con el 98.45% de las actas procesadas sin sumar los votos en actas inconsistentes o en actas de VMRE



Fuente: IFE



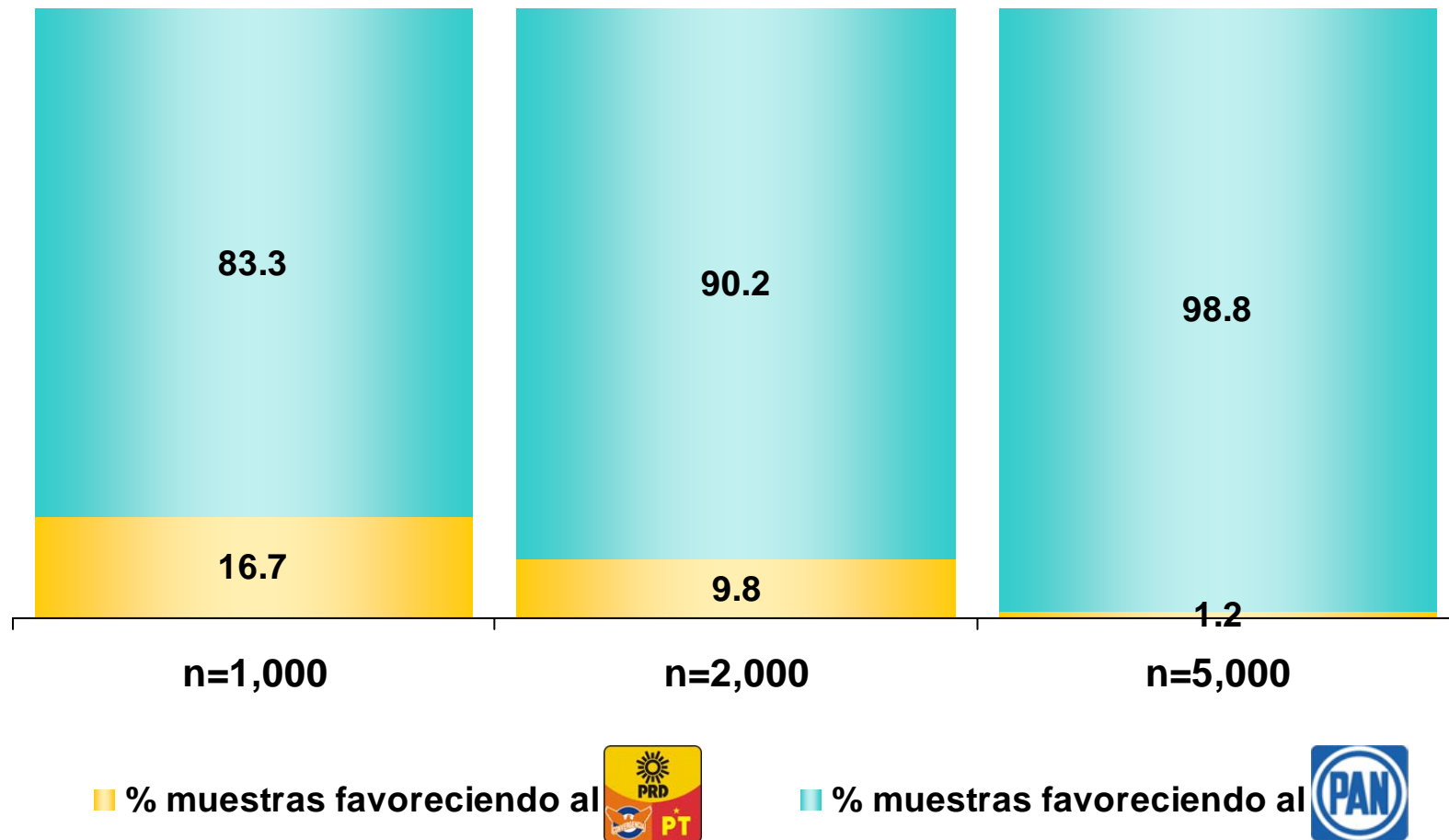
Diseño Aleatorio Simple

Ver características de los diseños en el Anexo 1



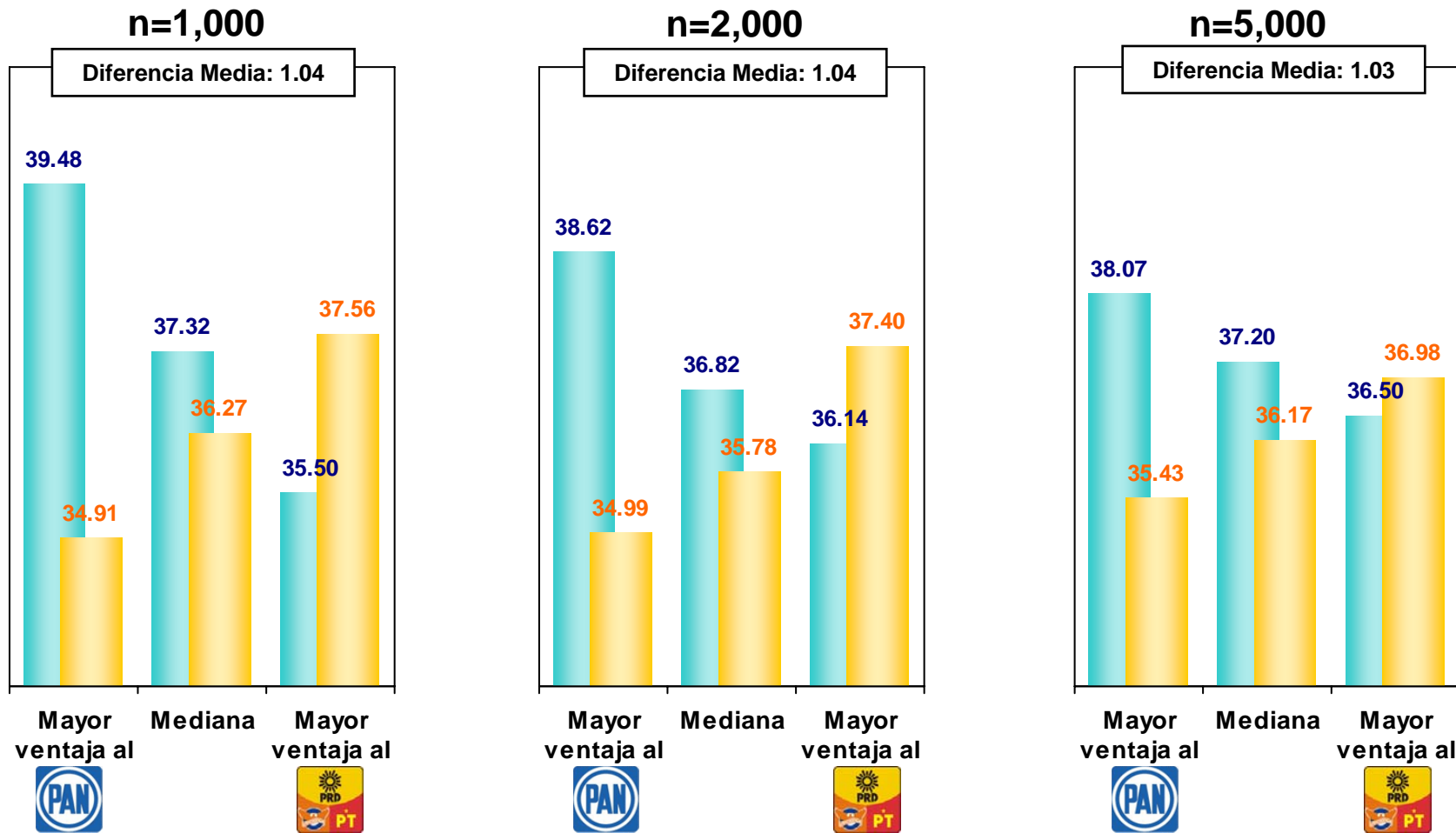
A mayores tamaños de muestra, se encuentran menos encuestas que favorezcan a la Alianza por el Bien de Todos

- Resultados de las 1,000 muestras a distintos tamaños



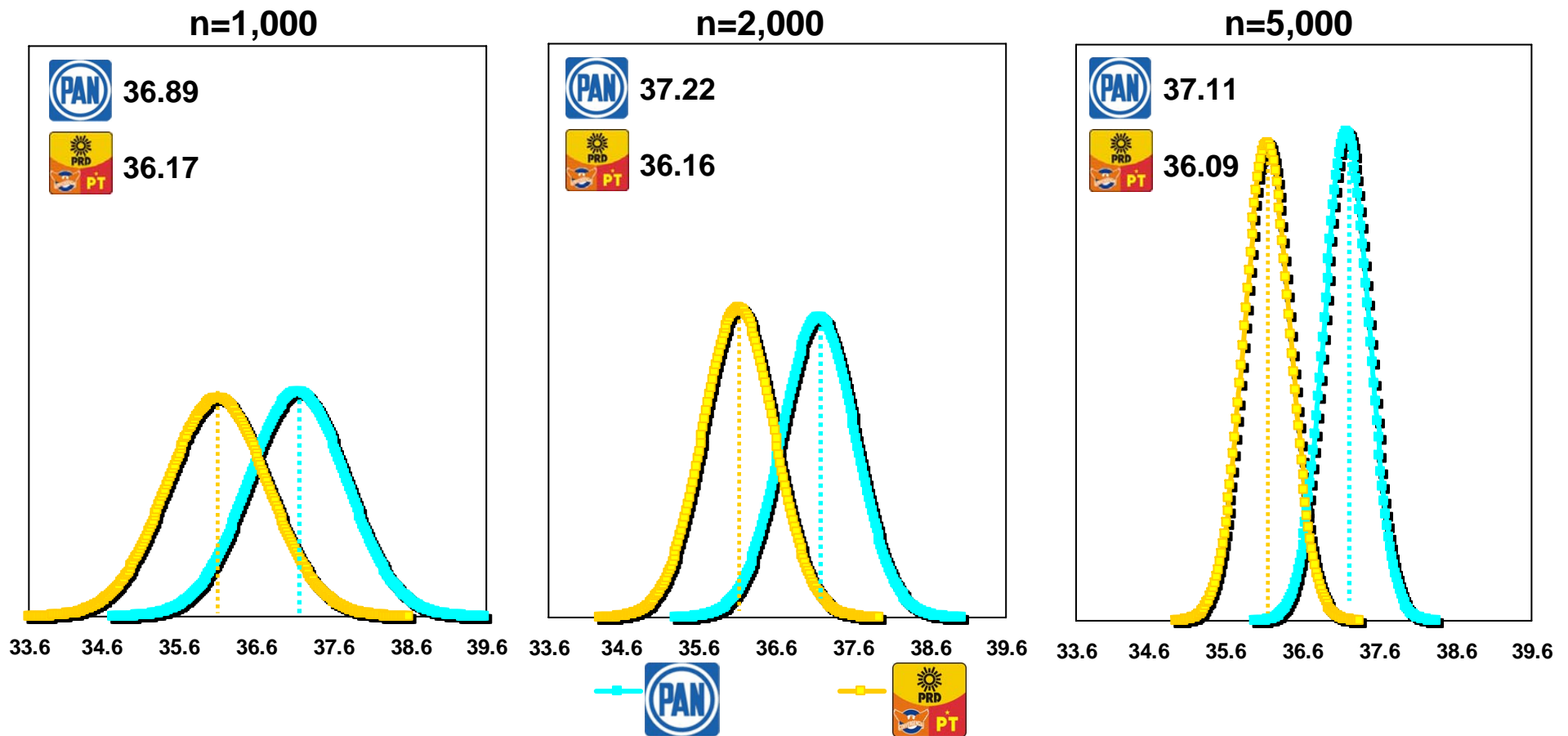
La ventaja de la Alianza por el Bien de Todos también se reduce a mayores tamaños de muestra

- Diferencias en resultados observados a distintos tamaños de muestra



Esto se explica debido a que a mayores tamaños de muestra, las áreas de coincidencia entre los resultados se reduce

- Curvas de Distribución para las medias de PAN y Alianza por el Bien de Todos



Distribuciones modeladas a partir de 1,000 muestras de cada tamaño, una vez corroborado que se cumple con los criterios de normalidad
Ver Método Bootstrap en el anexo

Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

La diferencia que se observa entre las cifras muestrales y las que se dieron al cierre del PREP es mínima

Resultados Preliminares
Preliminary Results
Résultats Préliminaires

Resumen Nacional
National Summary
Résumé National

Última actualización:
Corte a las 20:00 GMT - 06:00
del lunes 3 de julio del 2006

President / Presidente / Président

						Candidatos NO Registrados	Votos Nulos	Actas Procesadas	Total de Actas	Participación Ciudadana
Total Nacional	14,027,214	8,318,886	13,624,506	384,317	1,085,966	281,145	827,317	128,771	130,788	58.90%
Porcentaje	36.38%	21.57%	35.34%	00.99%	02.81%	00.72%	02.14%	98.45%		

Media de Mil muestras

n=5,000	37.15	22.05	36.12	1.02	2.88	0.79
---------	-------	-------	-------	------	------	------

1.04

1.03

Una de las posibles causas para que difiera el nivel puede darse en el método de muestreo (Aleatorio Simple) ya que para este tipo de casos es ideal asegurar representatividad y proporcionalidad con métodos estratificados

Fuente: IFE



Elección 2006

Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

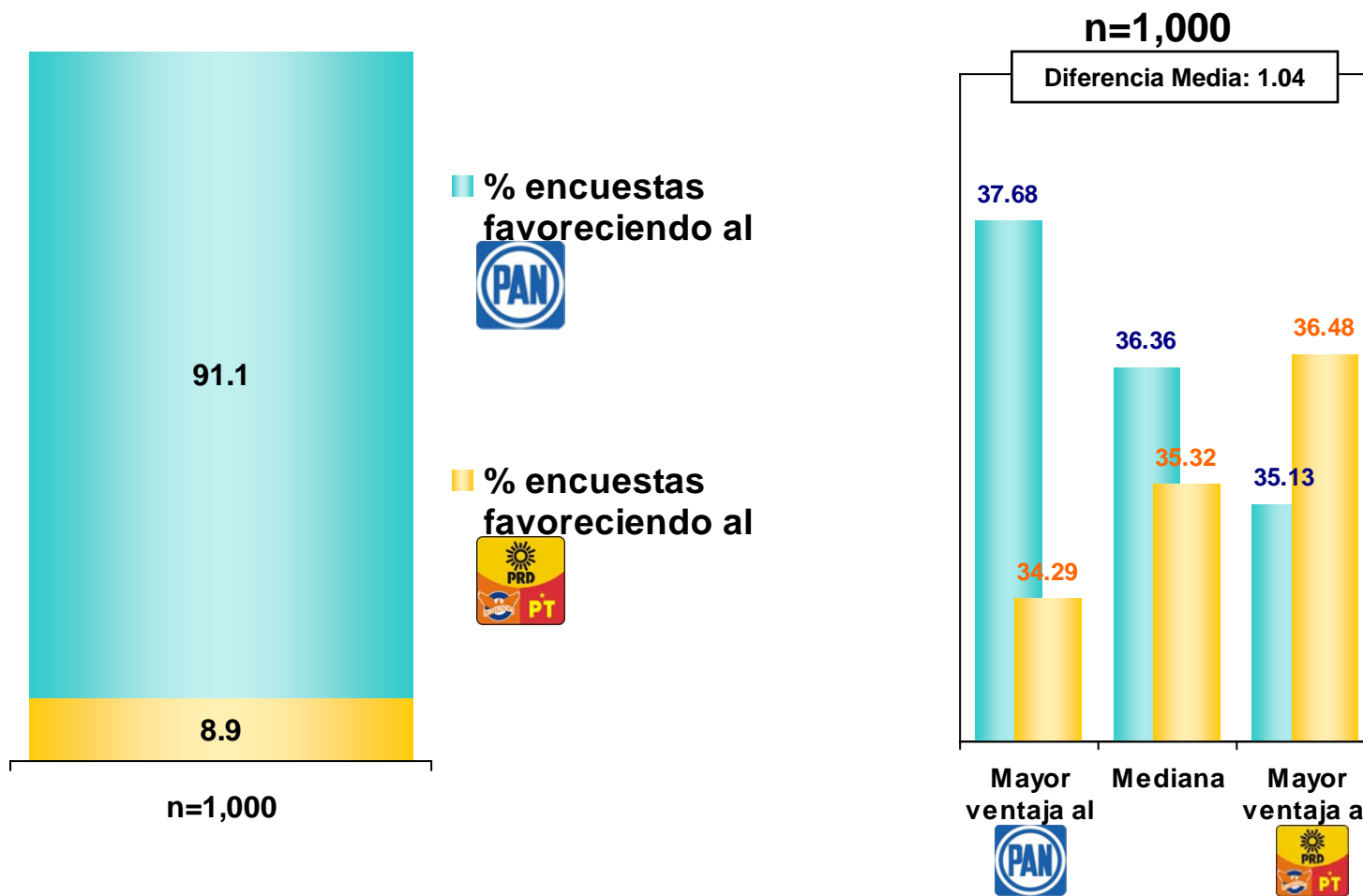


Diseño Estratificado




Con un diseño estratificado de muestras tamaño 1,000, se logran resultados más cercanos a los que se obtienen con el aleatorio simple de tamaño 2,000

- Resultados observados en mil muestras estratificadas de tamaño 1,000



Y el nivel de los resultados se vuelve más preciso




IFE
INSTITUTO FEDERAL ELECTORAL
ELECCIONES 2006

Voto en el extranjero

Presidente
Senadores
Diputados

CONSULTA EL RESULTADO DE TU CASILLA

Entidad Sección



PREP
PROGRAMA DE RESULTADOS
ELECTORALES PRELIMINARES
2006

Resultados Preliminares

Preliminary Results
Résultats Préliminaires






Resumen Nacional

National Summary
Résumé National

Última actualización:

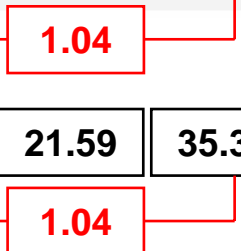
Corte a las 20:00 GMT - 06:00
del lunes 3 de julio del 2006

President / Presidente / Président

						Candidatos NO Registrados	Votos Nulos	Actas Procesadas	Total de Actas	Participación Ciudadana
Total Nacional	14,027,214	8,318,886	13,624,506	384,317	1,085,966	281,145	827,317	128,771	130,788	58.90%
Porcentaje	36.38%	21.57%	35.34%	00.99%	02.81%	00.72%	02.14%	98.45%		

Media de Mil muestras

n=1,000	36.38	21.59	35.34	1.00	2.82	2.87
---------	-------	-------	-------	------	------	------



Fuente: IFE



Elección 2006

Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

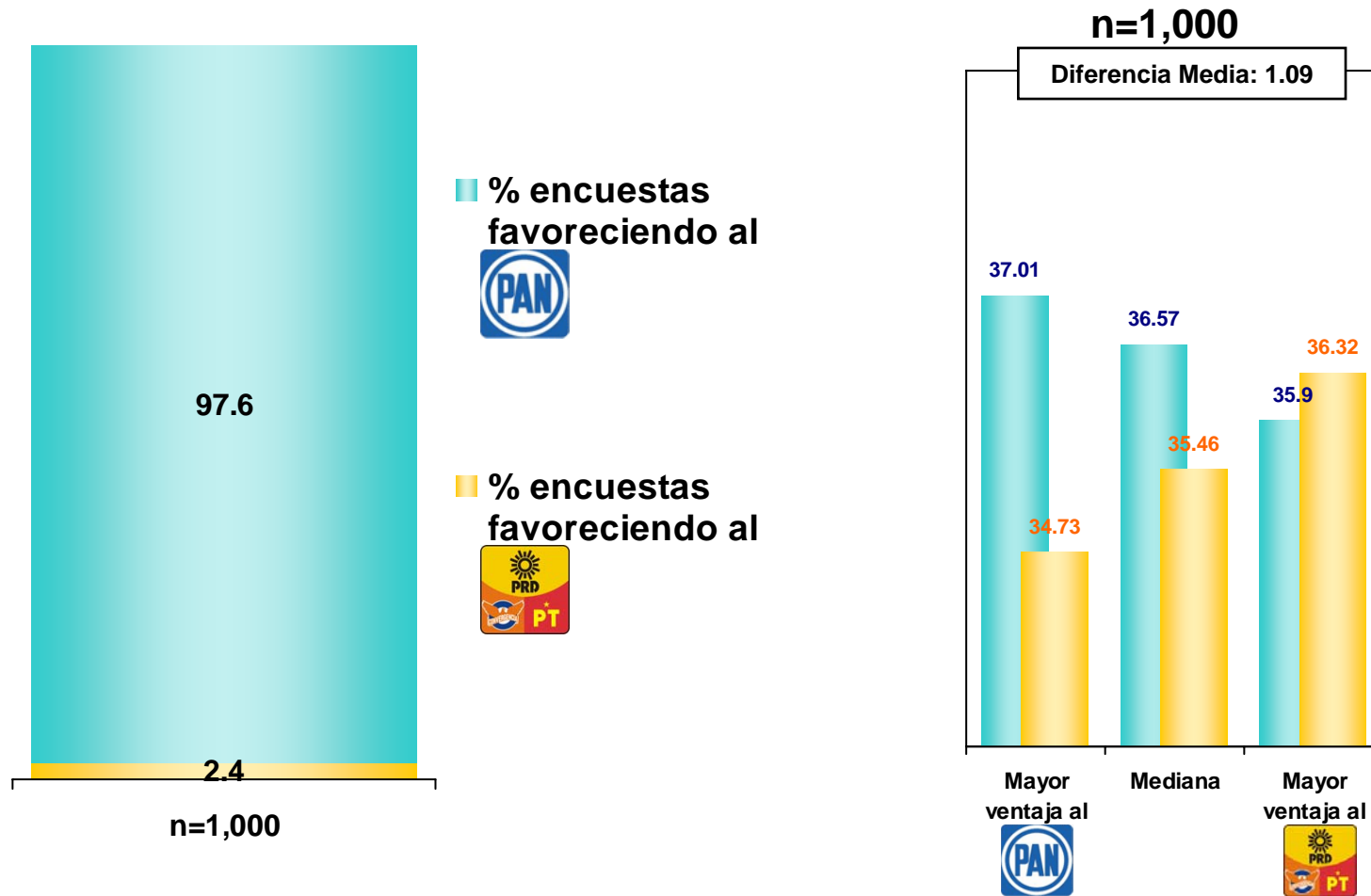


Diseño Sistemático sobre Lista Ordenada (Estratificación Implícita)



Con un diseño estratificado, se logran resultados más cercanos con muestras considerablemente menores

- Resultados observados en 40 muestras estratificadas de tamaño 1,000



Además de apuntar en la misma dirección, el nivel de los resultados se vuelve más preciso

Resultados Preliminares
 Preliminary Results
 Résultats Préliminaires

Resumen Nacional
 National Summary
 Résumé National

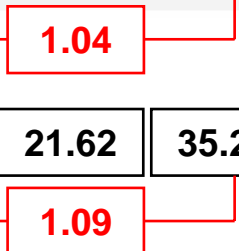
Última actualización:
 Corte a las 20:00 GMT - 06:00
 del lunes 3 de julio del 2006

President / Presidente / Président

						Candidatos NO Registrados	Votos Nulos	Actas Procesadas	Total de Actas	Participación Ciudadana
Total Nacional	14,027,214	8,318,886	13,624,506	384,317	1,085,966	281,145	827,317	128,771	130,788	58.90%
Porcentaje	36.38%	21.57%	35.34%	00.99%	02.81%	00.72%	02.14%	98.45%		

Media de 40 muestras

n=1,000	36.36	21.62	35.29	1.01	2.83	2.87
---------	-------	-------	-------	------	------	------



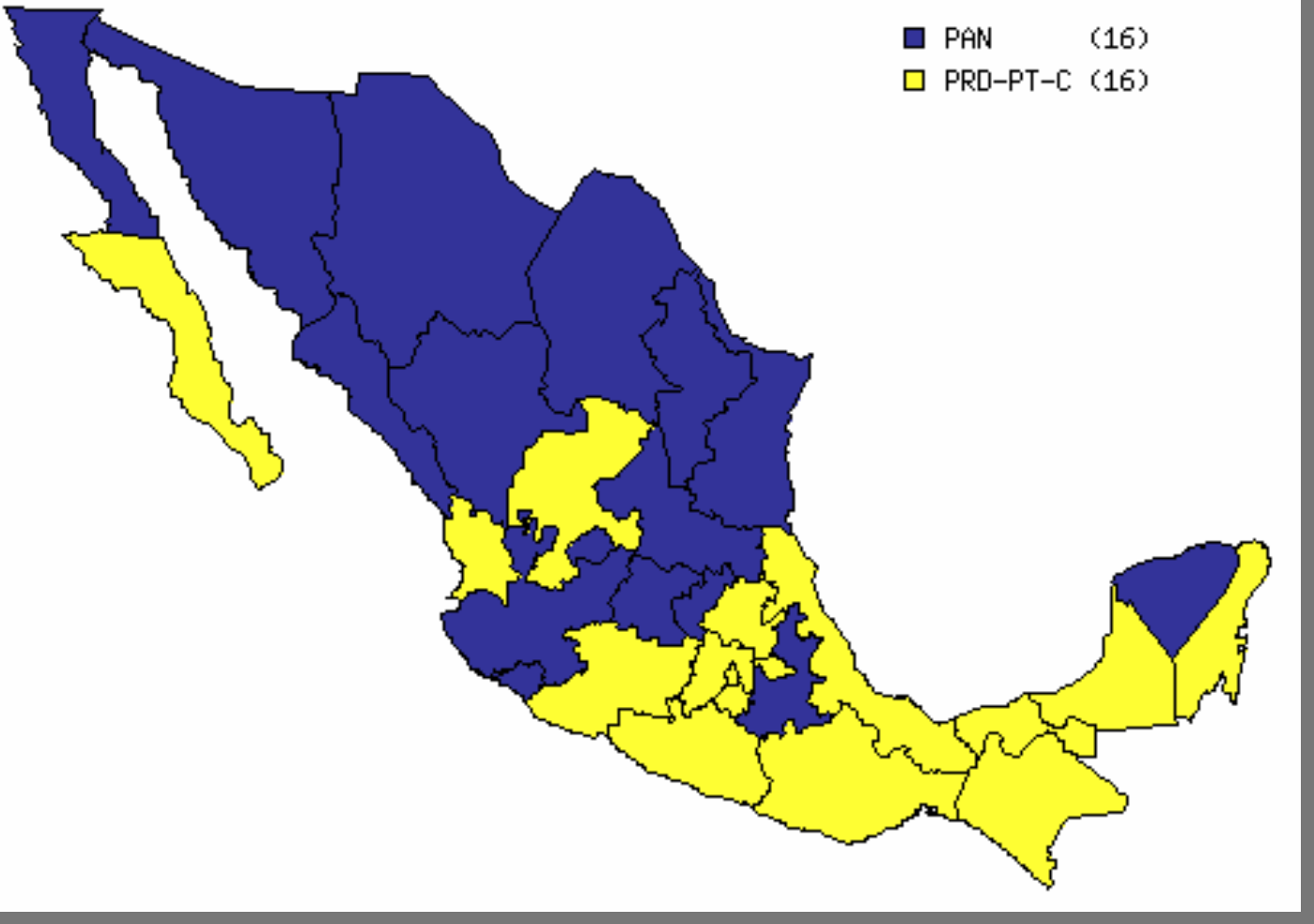
Fuente: IFE



Elección 2006

Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

Los diseños estratificados son más precisos cuando existen diferencias geográficas claras



Fuente: IFE



Elección 2006

Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

Sumario

- Con muestras de tamaño 1,000 y un diseño aleatorio simple, es posible obtener una diferencia similar a la que arrojó el PREP, aunque el nivel de variabilidad es tan amplio que no es posible precisar el resultado
 - A este tamaño de muestra 17% de las muestras daban como ganador a la Alianza por el Bien de Todos
- Con muestras mayores (n=5,000) la variabilidad se reduce, de tal manera, que la probabilidad de una victoria para la Alianza por el Bien de Todos se es muy baja
- Sin embargo, al utilizar un diseño estratificado, la variabilidad se reduce aún más e incluso con una muestra de tamaño 1,000, no se observan casos en los que se obtenga el triunfo por parte de la Alianza por el Bien de Todos

[Regresar a Reflexiones](#)



**¿Realizar cortes en el PREP implica
analizar muestras representativas?
Pruebas de Bondad de Ajuste**



Metodología

- Con base en las cifras de las actas de las casillas que se recibieron, se procesaron y se contabilizaron en el PREP 2006, se realizaron cortes acumulados cada 20 minutos, a partir de las 19:00 hrs del 2 de julio de 2006, en los que se calculaban:
 - Porcentaje de actas computadas, en relación al universo total de actas
 - Porcentaje de votos computados, en relación al universo total de actas
 - Porcentaje de los votos computados a favor de cada candidato presidencial, en relación al universo total de actas
 - La proporción acumulada, a cada corte, de los votos obtenidos a favor de cada uno de los candidatos
- Con estos resultados, se realizó un análisis descriptivo de las tendencias mostradas por el PREP en el que se determinaron cortes de interés para llevar a cabo análisis más detallados y estadísticamente formales



Elección de Cortes Interesantes



El análisis contempla los resultados obtenidos de la contabilización del 98.45% de las actas totales

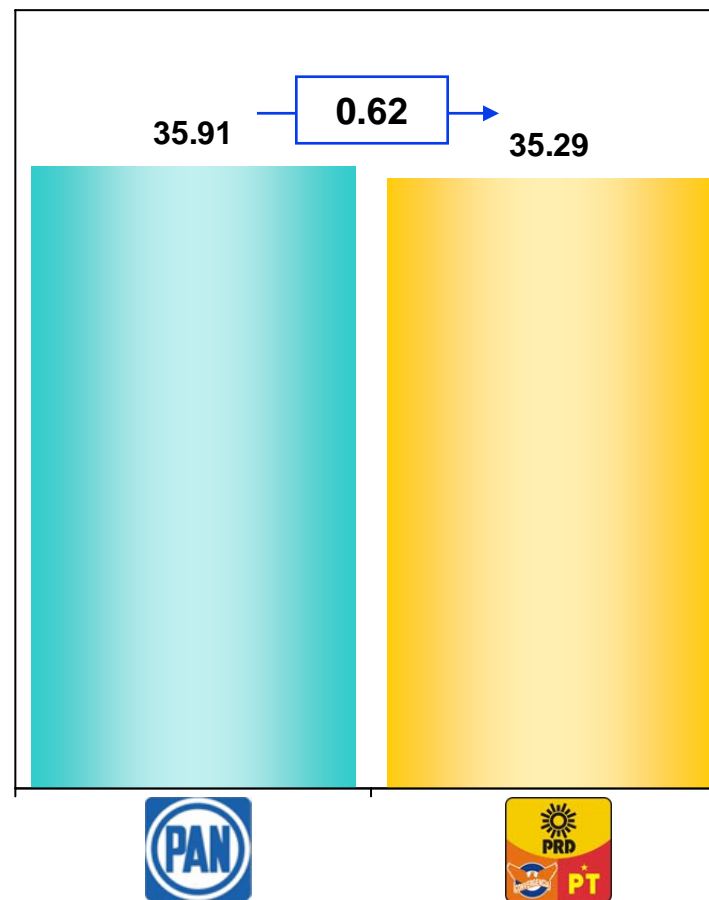
- Universo de Actas sobre las que se Realizó el Análisis

Actas Contabilizadas:	117,287
Actas Inconsistentes:	11,184
VMRE:	300
Total Actas Procesadas:	128,771

Total Nacional + VMRE: 130,788

98.45% de las Actas Totales*

- Resultados por Partido contemplado en el Total Actas Procesadas

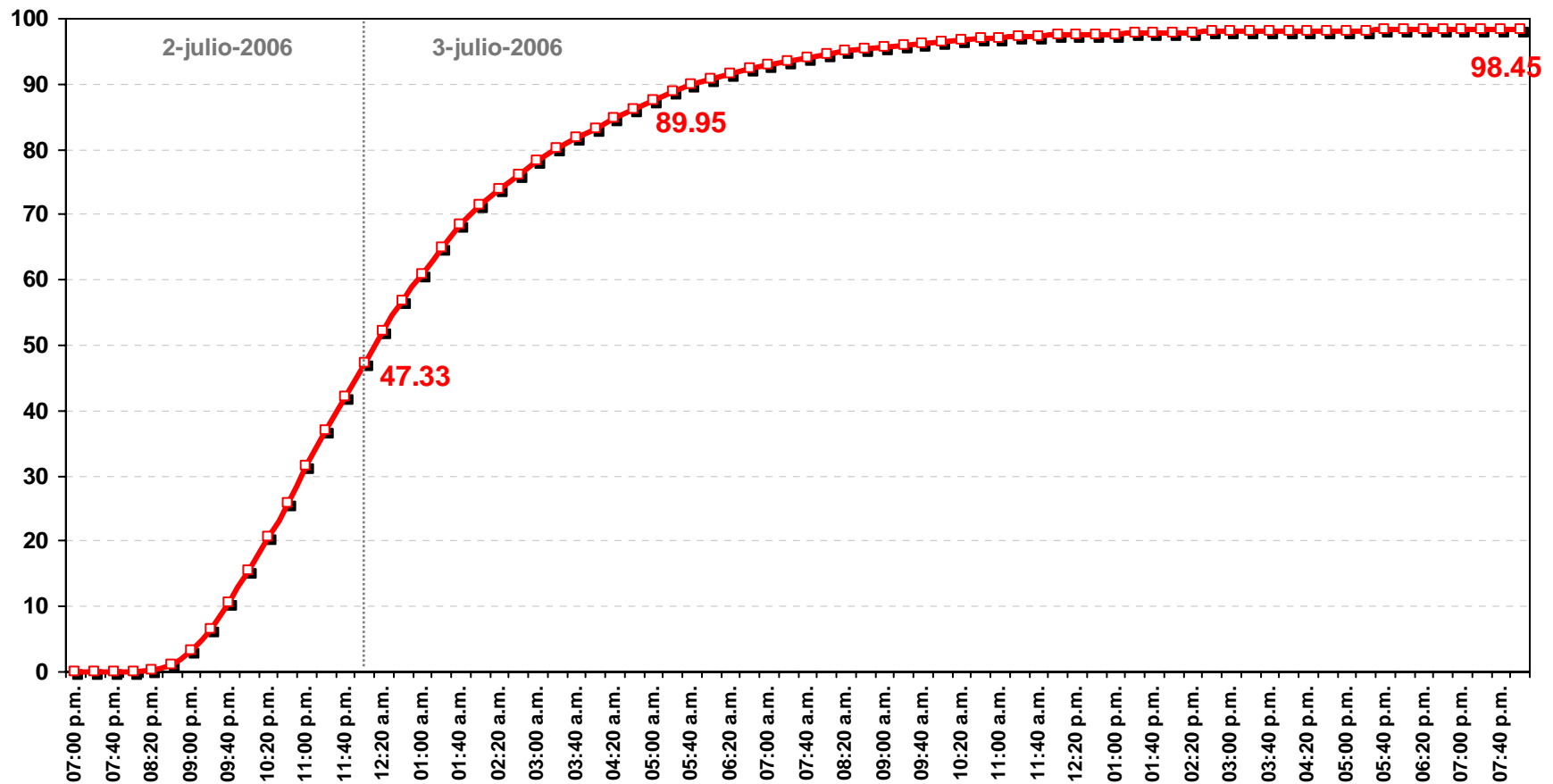


* El 1.55% restante corresponde a casillas sin

Fuente: IFE actas y a actas no recibidas

A las 00:00 hrs. del 3 de julio de 2006, se había procesado el 47.33% de las actas y para las 20:00 hrs se tenía el 98.45%

- % Acumulado de Actas Procesadas por Hora de Registro en el PREP (Cortes de 20 min.)



Fuente: IFE

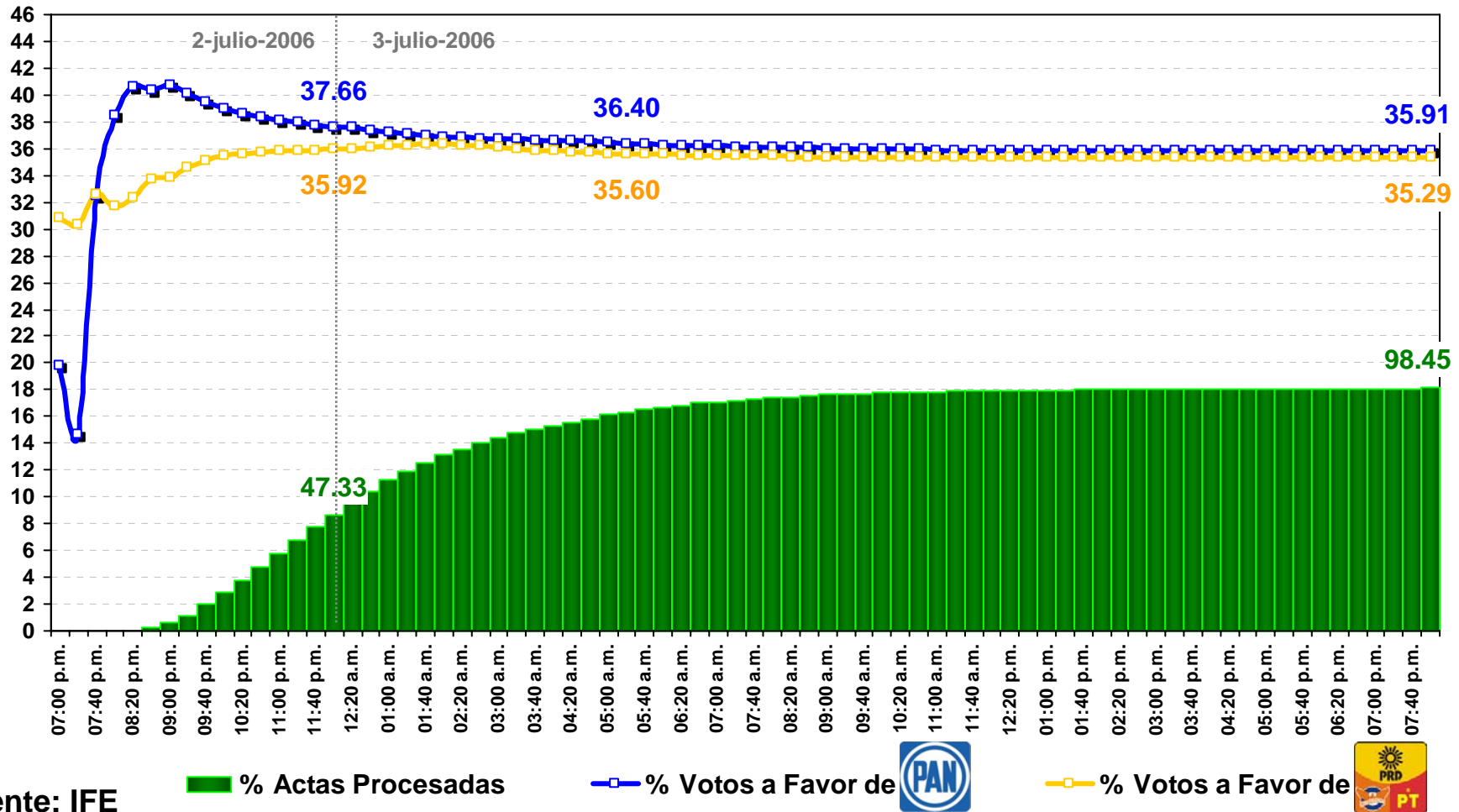
Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business



Elección 2006

A esa hora, la proporción de votos favorables a Calderón era mayor a la de López Obrador, manteniéndose por arriba hasta el cierre

- % Votos Obtenidos a Favor de los Dos Principales Candidatos a la Presidencia del País



Fuente: IFE

■ % Actas Procesadas
 —□— % Votos a Favor de PAN
 —□— % Votos a Favor de PRD PT

Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business



Elección 2006

Reflexiones a Partir de los Resultados del PREP

- Con base en los resultados procesados por el PREP, se observa que:
 - Previo al corte acumulado de las 19:40 hrs., del 2 de julio de 2006, la proporción de votos a favor de López Obrador es superior a la obtenida por Calderón
 - Posteriormente a este corte, se observa una ventaja para Calderón y aunque la diferencia entre las proporciones obtenidas a favor de cada uno de los candidatos se hace menor con el tiempo, en ningún otro corte acumulado se observa que este comportamiento se revierta
- De lo anterior, se desprenden las siguientes reflexiones:
 - ¿Cuál sería una causa probable para observar una ventaja de López Obrador hasta antes de las 19:40 hrs. que no se volvió a observar?
 - ¿A qué se le podría atribuir el hecho de que la ventaja de Calderón vs. López Obrador se reduzca pero que no se vuelva a dar un cruce entre las proporciones?

Razonamientos ante las Reflexiones

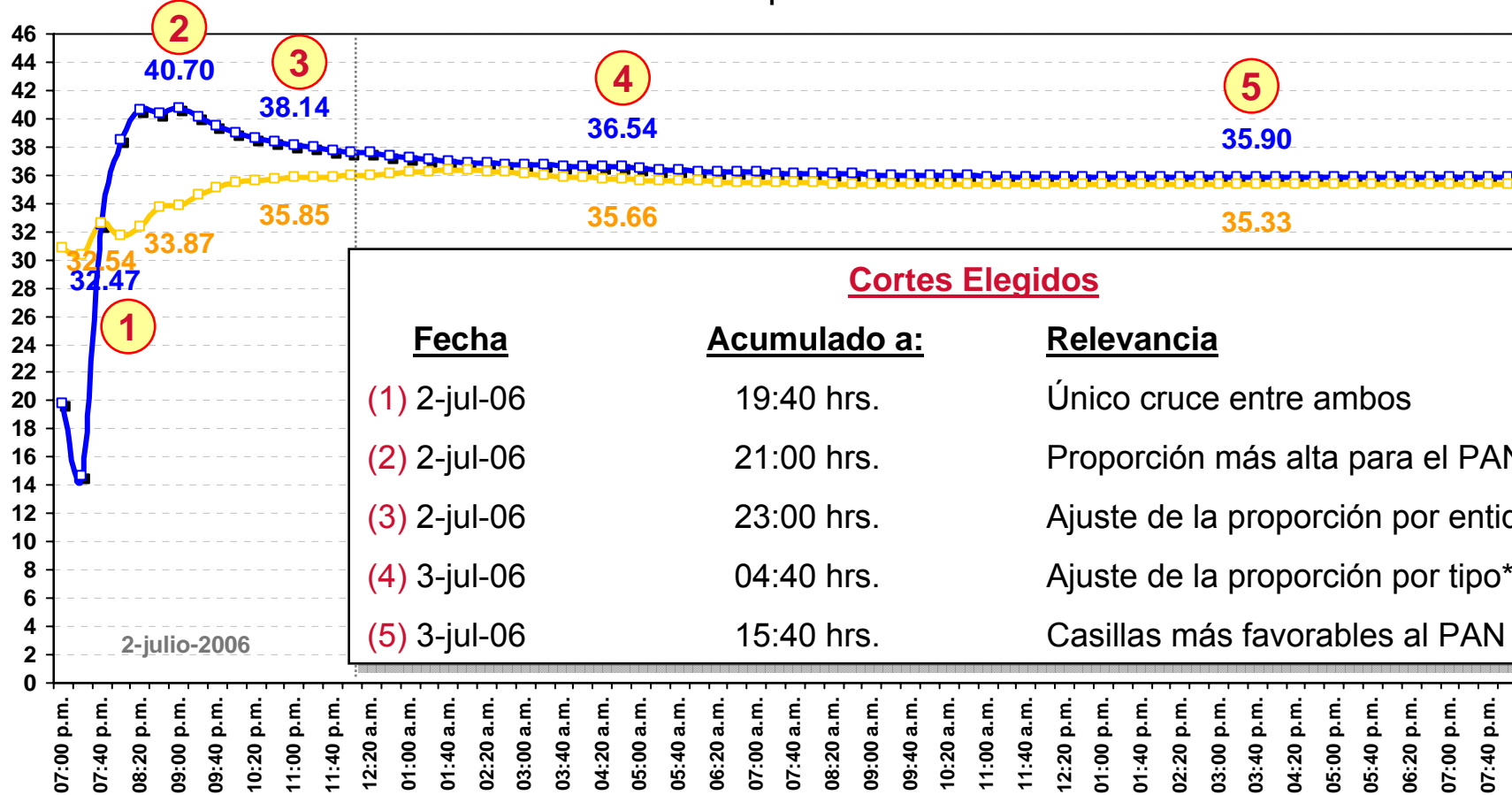
- **¿Cuál sería una causa probable para observar una ventaja de López Obrador hasta antes de las 19:40 hrs. que no se volvió a observar?**
 - De acuerdo al comportamiento observado, para las proporciones de cada candidato hasta antes de las 19:40 hrs. y, en contraste, con el comportamiento posterior, se puede establecer la hipótesis de que los votos procesados hasta ese momento no conformaban una muestra que representara correctamente al universo total de votos
 - Para poder suponer que una muestra sea representativa, entre otras cosas, esperaríamos como un mínimo requerimiento que la estructura de los votos procesados, para cada entidad federativa, mostrara una estructura similar a la que se da en el universo de referencia
 - Adicionalmente al supuesto anterior, se le podrían agregar criterios más estrictos, como que la proporción de votos procesados en casillas urbanas y no urbanas también sea similar a la que se da en el universo de referencia

Razonamientos ante las Reflexiones

- **¿A qué se le podría atribuir el hecho de que la ventaja de Calderón vs. López Obrador se reduzca pero que no se vuelva a dar un cruce entre las proporciones?**
 - También con base en el comportamiento observado se pueden establecer hipótesis similares a la de la reflexión anterior
 - La primer hipótesis establecería que al momento en el que la proporción de votos a favor de Calderón se encuentra alrededor del 40%, la muestra pudiera no ser representativa del universo
 - Una segunda hipótesis establece que en la medida en que se procese un mayor porcentaje de actas, la representatividad de este acumulado, con respecto al universo es cada vez mayor. Al momento en que dicha representatividad se alcance, el comportamiento en las proporciones de los votos obtenidos por los candidatos tenderá cada vez más hacia el resultado final, volviendo muy poco probable que se dé un cruce entre las mismas

Para dar tratamiento a estas hipótesis se han elegido 5 cortes a lo largo del tiempo en que fueron registrándose los resultados

- % Votos Obtenidos a Favor de los Dos Principales Candidatos a la Presidencia del País



—■ % Votos a Favor de 

—■ % Votos a Favor de 

* Casilla Urbana o Casilla No Urbana

Fuente: IFE



Elección 2006

Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business



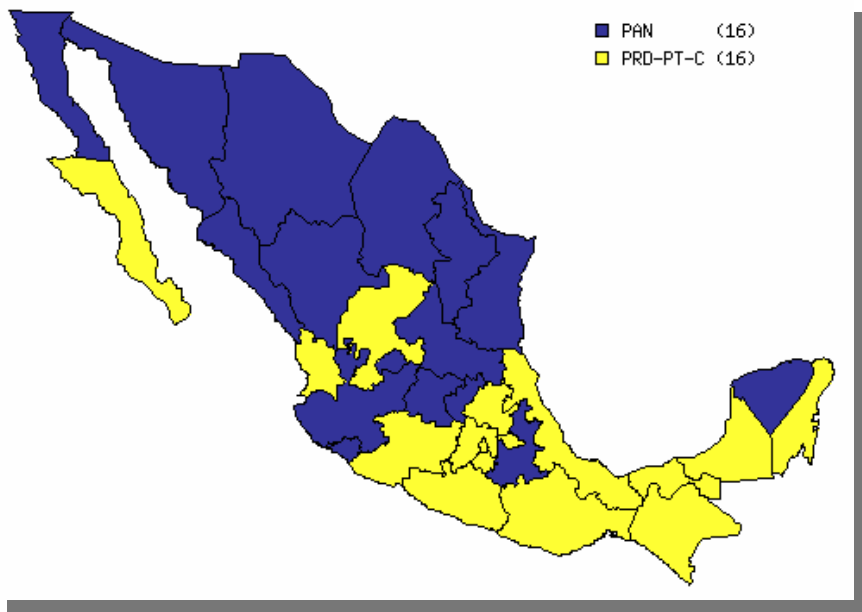
Pruebas de Bondad de Ajuste Ji-cuadrada

Ver características de la prueba en el Anexo 3



En estas elecciones se observó una polarización en las preferencias de los votantes a nivel entidad federativa

- Mapa de Partidos Triunfantes en cada Entidad Federativa



- % de la Lista Nominal Distribuida a Nivel Entidad Federativa

Entidad	% Lista Nominal
Aguascalientes	1.00
Baja California	2.81
Baja California Sur	0.46
Campeche	0.66
Coahuila	2.40
Colima	0.58
Chiapas	3.66
Chihuahua	3.07
Distrito Federal	10.03
Durango	1.45
Guanajuato	4.82
Guerrero	2.97
Hidalgo	2.29
Jalisco	6.63
México	12.81
Michoacán	4.14

Entidad	% Lista Nominal
Morelos	1.66
Nayarit	0.96
Nuevo León	4.14
Oaxaca	3.27
Puebla	4.85
Querétaro	1.45
Quintana Roo	0.94
San Luis Potosí	2.25
Sinaloa	2.43
Sonora	2.34
Tabasco	1.87
Tamaulipas	3.11
Tlaxcala	1.00
Veracruz	6.90
Yucatán	1.64
Zacatecas	1.40

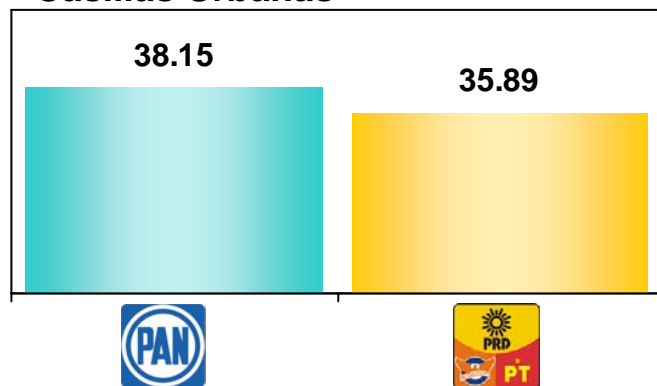
Ante un comportamiento de este tipo, para que una muestra de votos sea representativa, al menos debe cumplir con las proporciones de votos que guarda el universo por entidad federativa

Fuente: IFE

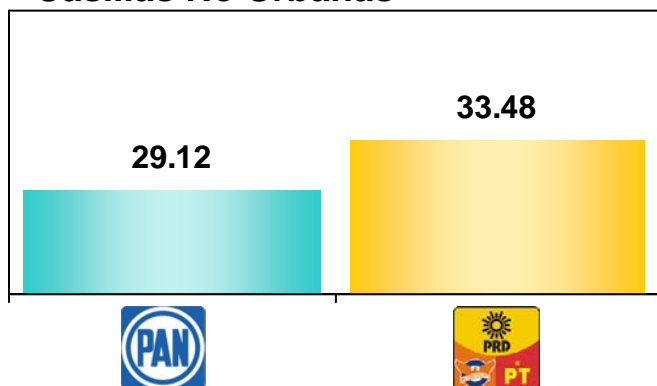
También se observa una diferencia en las preferencias para cada tipo de casilla

- % Votos a Favor de los Dos Principales Candidatos por Tipo de Casilla

Casillas Urbanas



Casillas No Urbanas



Fuente: IFE

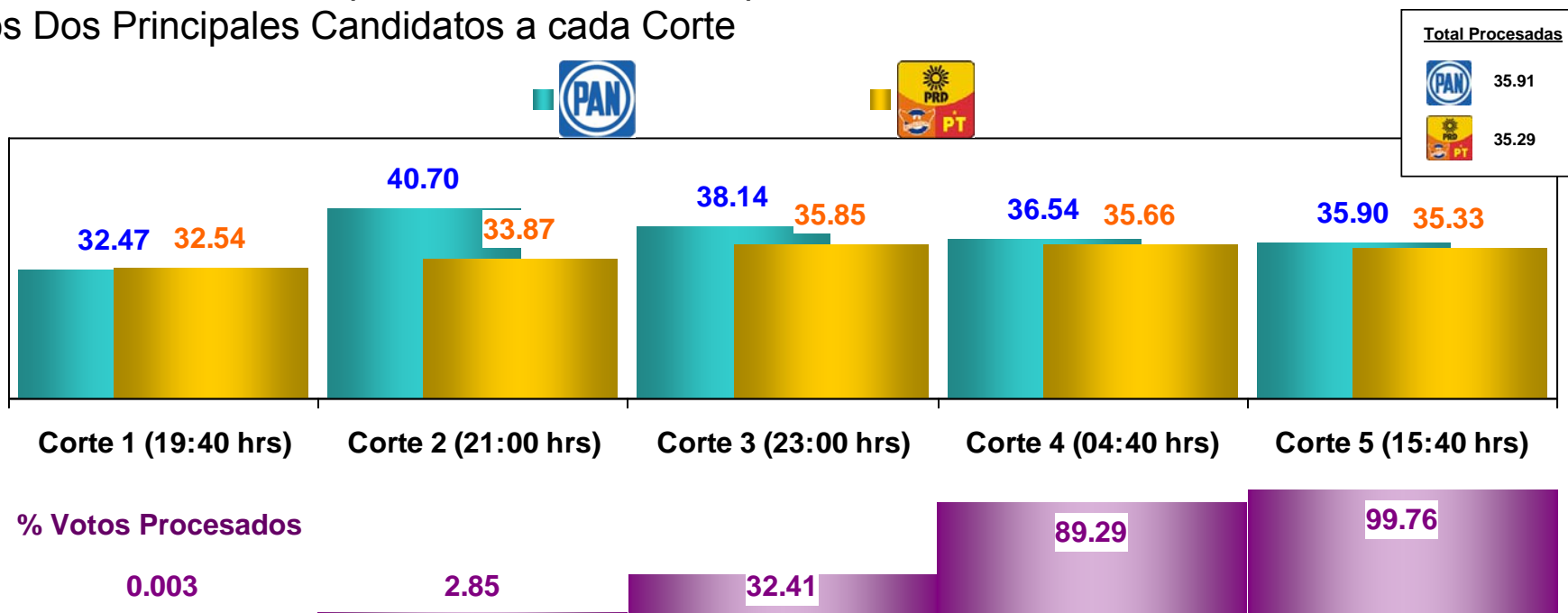
- % de la Lista Nominal Distribuida en cada Tipo de Casilla

Tipo de Casilla	% Lista Nominal
Urbana	73.24
No Urbana	26.76

Debido a la diferencia en las preferencias, en la medida en la que una muestra de votos se apegue más a las proporciones del universo, se logrará una mayor representatividad del resultado final

Al realizar pruebas de bondad, a diferentes cortes, se determina que hasta las 23:00 hrs. del 2 de julio se logran las proporciones esperadas por entidad

- Resultados de las Proporciones Obtenidas a partir de los Votos Procesados a Favor de los Dos Principales Candidatos a cada Corte



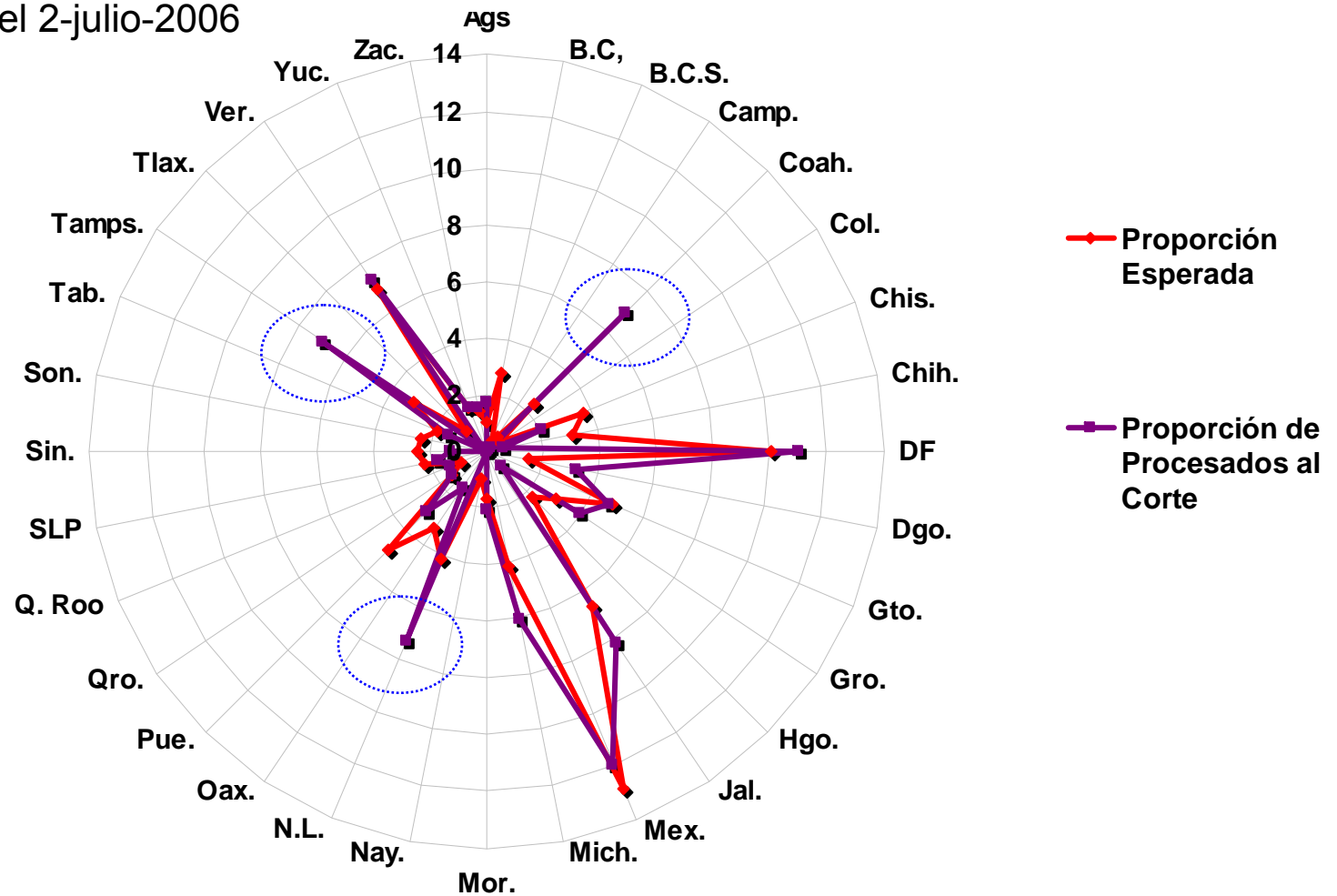
Prueba de Bondad de Ajuste ¿Las proporciones de votos procesados cumplen con lo esperado en:

Entidades?	NO	NO	SÍ	SÍ	SÍ
Tipo de Casilla?	NO	NO	NO	SÍ	SÍ

Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

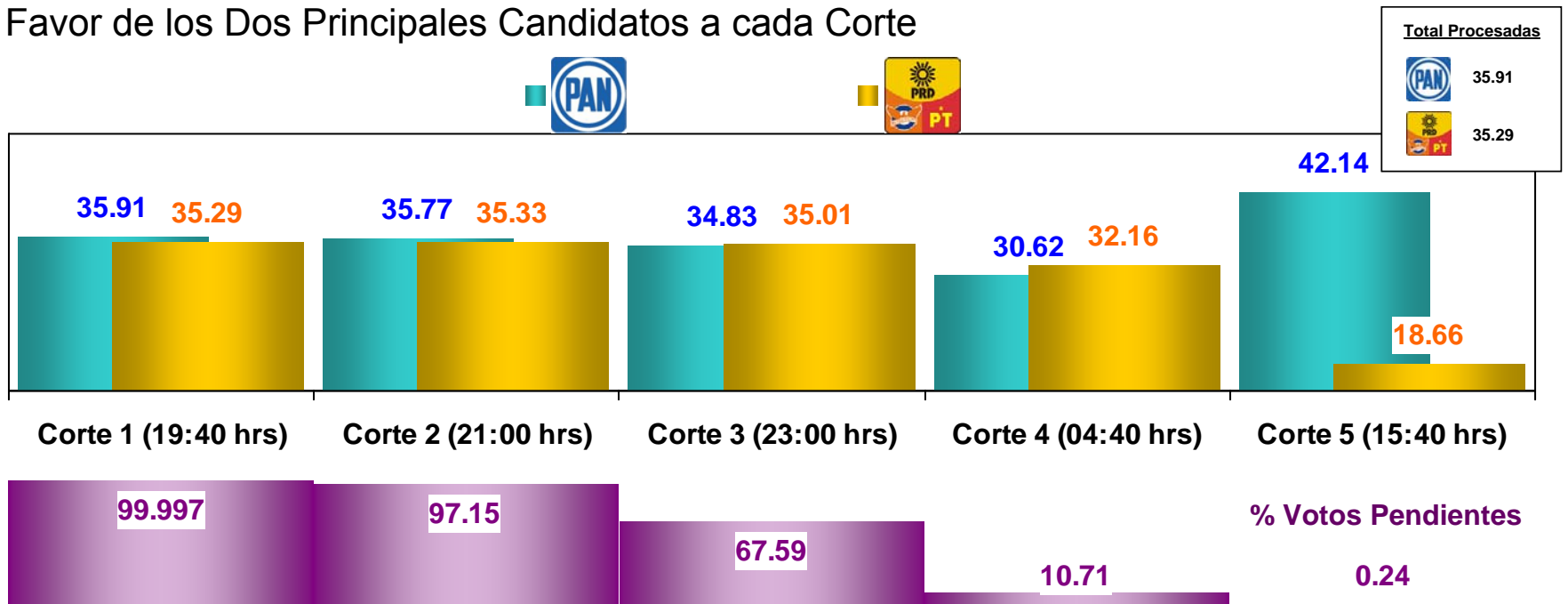
Al corte de las 21:00 hrs. se dio uno de los niveles más altos para el PAN y al analizarlo se encuentra que 3 entidades con proporciones mayores a las esperadas son del Norte

- Proporciones Observadas de Votos Procesados por Entidad Federativa al Corte de las 21:00 hrs. del 2-julio-2006



Por otro lado, también se encuentra que para un porcentaje menor al 1% de los votos pendientes de procesar se da una mayor preferencia por el PAN

- Resultados de las Proporciones Obtenidas a partir de los Votos Pendientes de Procesar a Favor de los Dos Principales Candidatos a cada Corte

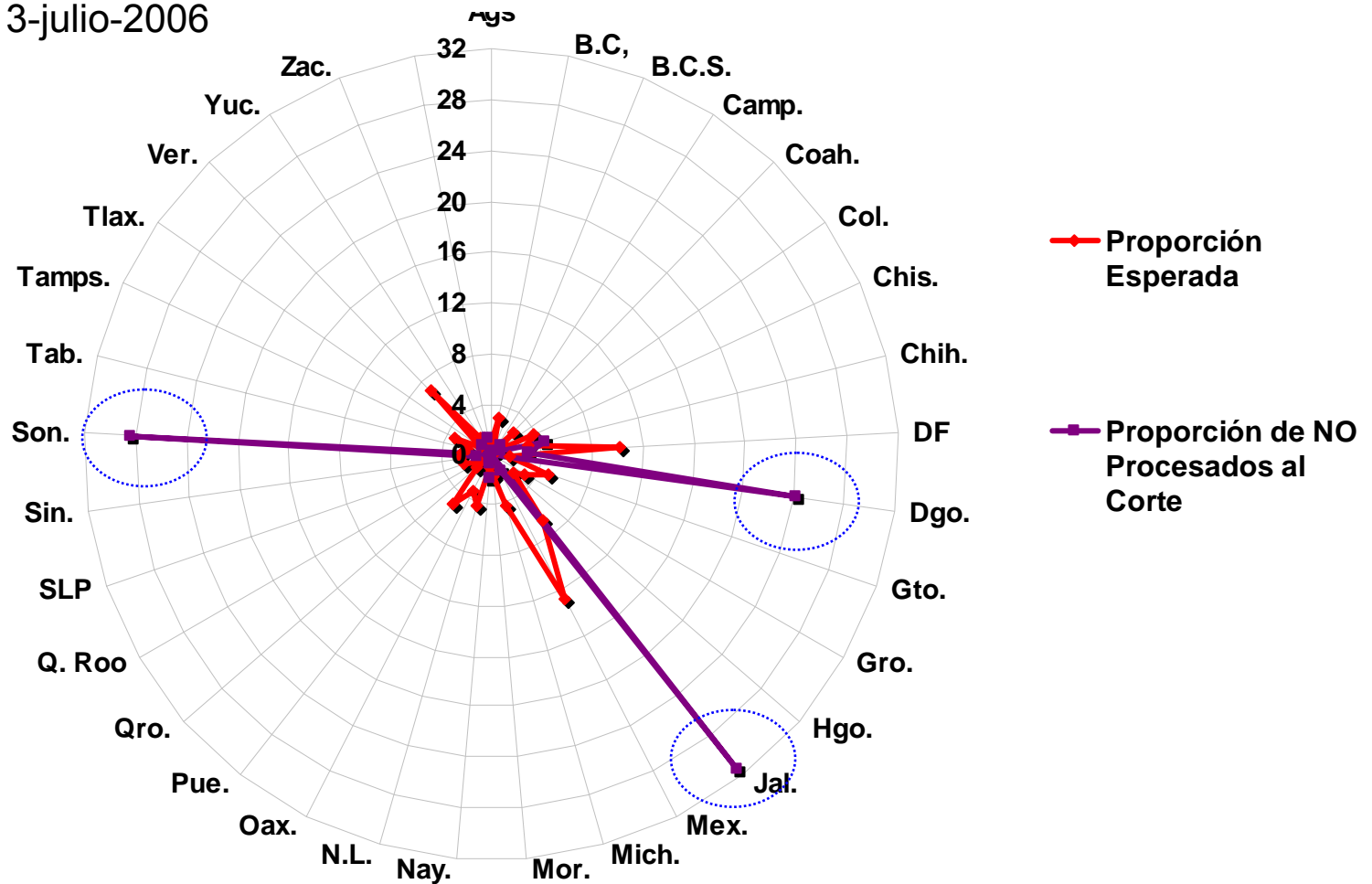


Prueba de Bondad de Ajuste ¿Las proporciones de votos pendientes cumplen con lo esperado en:

Entidades?	SÍ	SÍ	SÍ	NO	NO
Tipo de Casilla?	SÍ	SÍ	SÍ	NO	NO

Observando las proporciones se vuelve a confirmar que las entidades que presentan proporciones mayores a las esperadas son del Norte del país

- Proporciones Observadas de Votos Pendientes por Entidad Federativa al Corte de las 15:40 hrs. del 3-julio-2006





Análisis para cada uno de los Cortes

Ver sustento del análisis en el Anexo 4





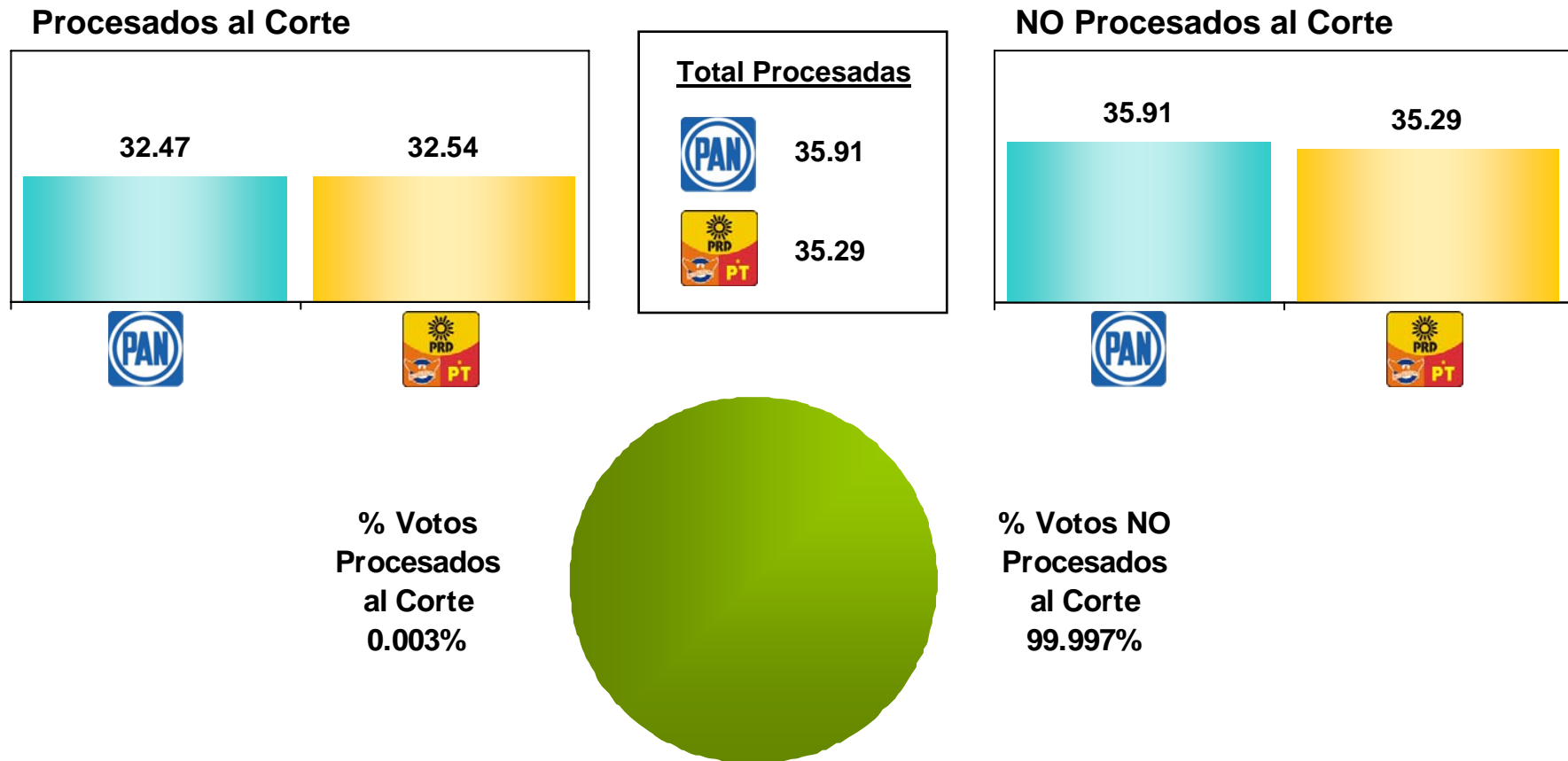
Corte 1. 19:40 hrs. del 2-julio-2006



Hasta el acumulado de las 19:40 hrs. la coalición Por el Bien de Todos mantenía una ventaja en los votos procesados, aunque estos no sumaban ni el 1% del total de votos

Corte 1. Acumulado a las 19:40 hrs del 2-julio-2006

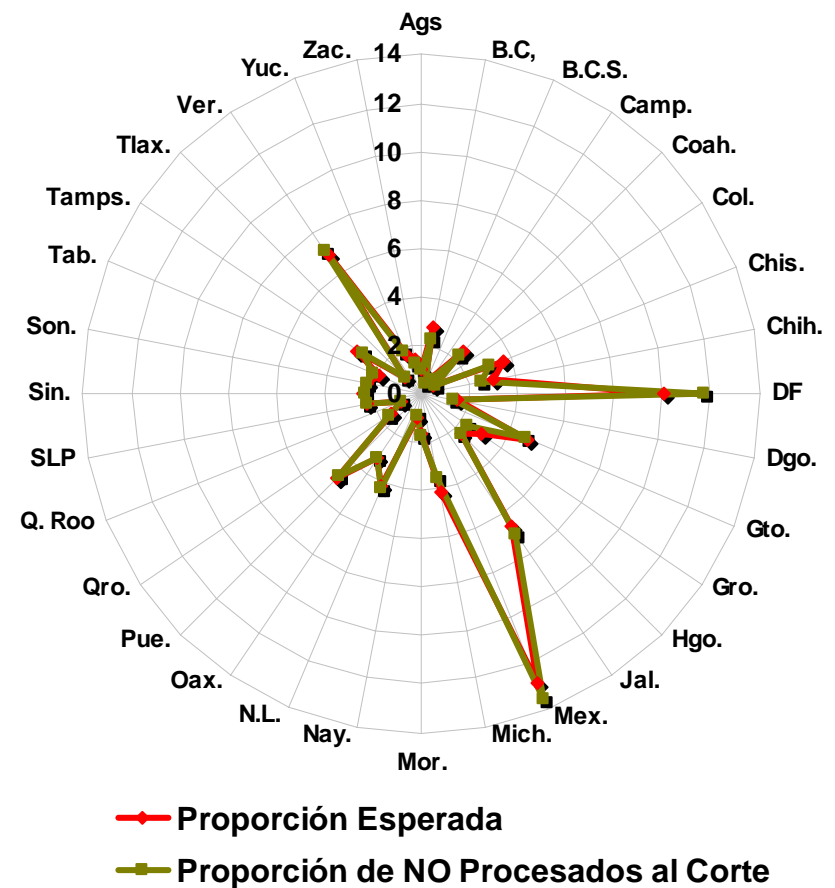
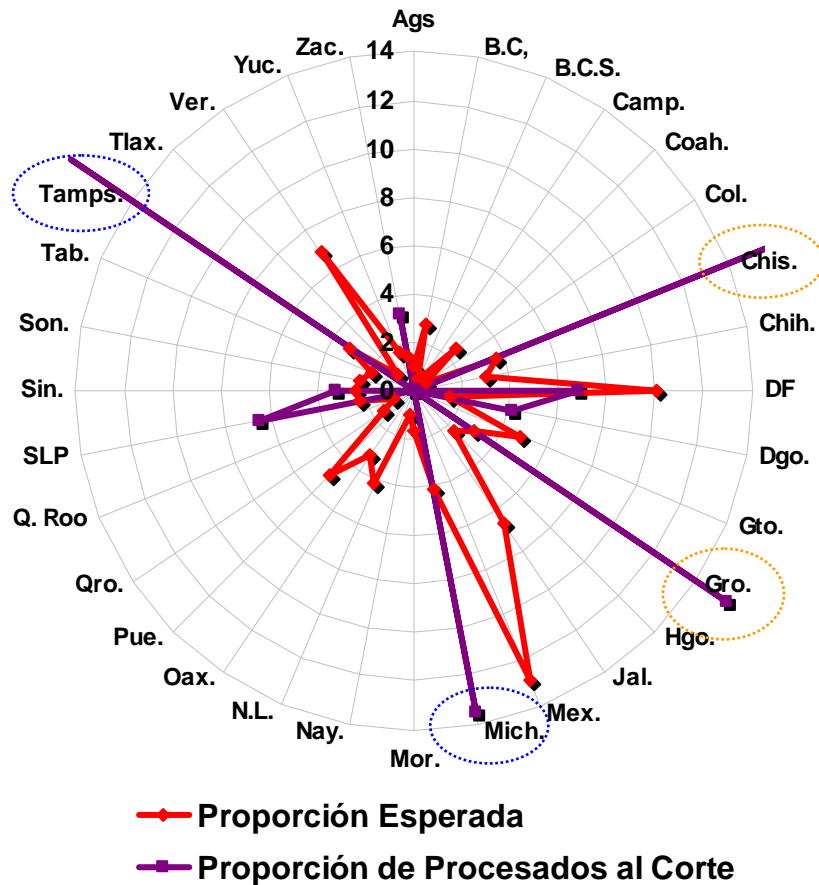
- Resultados de las Proporciones de Votos a Favor de los Dos Principales Candidatos



Chiapas, Guerrero, Tamaulipas y Michoacán mostraban proporciones por arriba de las esperadas

Corte 1. Acumulado a las 19:40 hrs del 2-julio-2006

- % de Votos por Entidad Federativa con Respecto al Total Nacional de Votos



Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

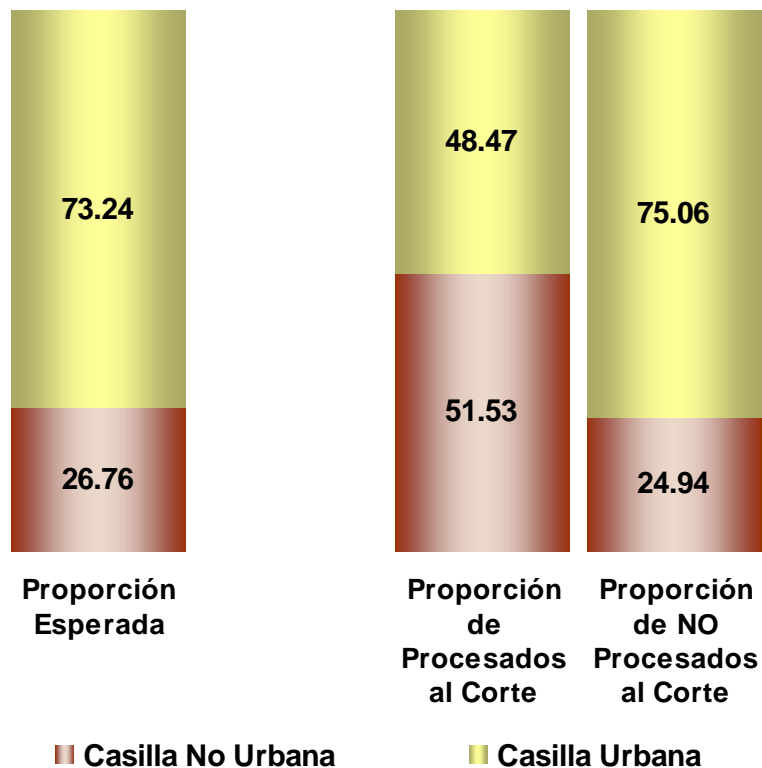


Elección 2006

Y las casillas No Urbanas también mostraban un mayor peso que el esperado, influyendo en que las pruebas de bondad de ajuste confirmen la falta de ajuste a lo esperado

Corte 1. Acumulado a las 19:40 hrs del 2-julio-2006

- % de Votos por Tipo de Casilla con Respecto al Total Nacional de Votos



- Resultados Obtenidos por la Prueba de Bondad de Ajuste (ji-cuadrada)

Para las 32 entidades federativas

¿Las proporciones son iguales?		X ²
Para los votos procesados al corte	NO	37.98 x10 ⁷
Para los votos NO procesados al corte	SÍ	1.057

Para los 2 tipos de casillas

¿Las proporciones son iguales?		X ²
Para los votos procesados al corte	NO	24.554
Para los votos NO procesados al corte	SÍ	0.177



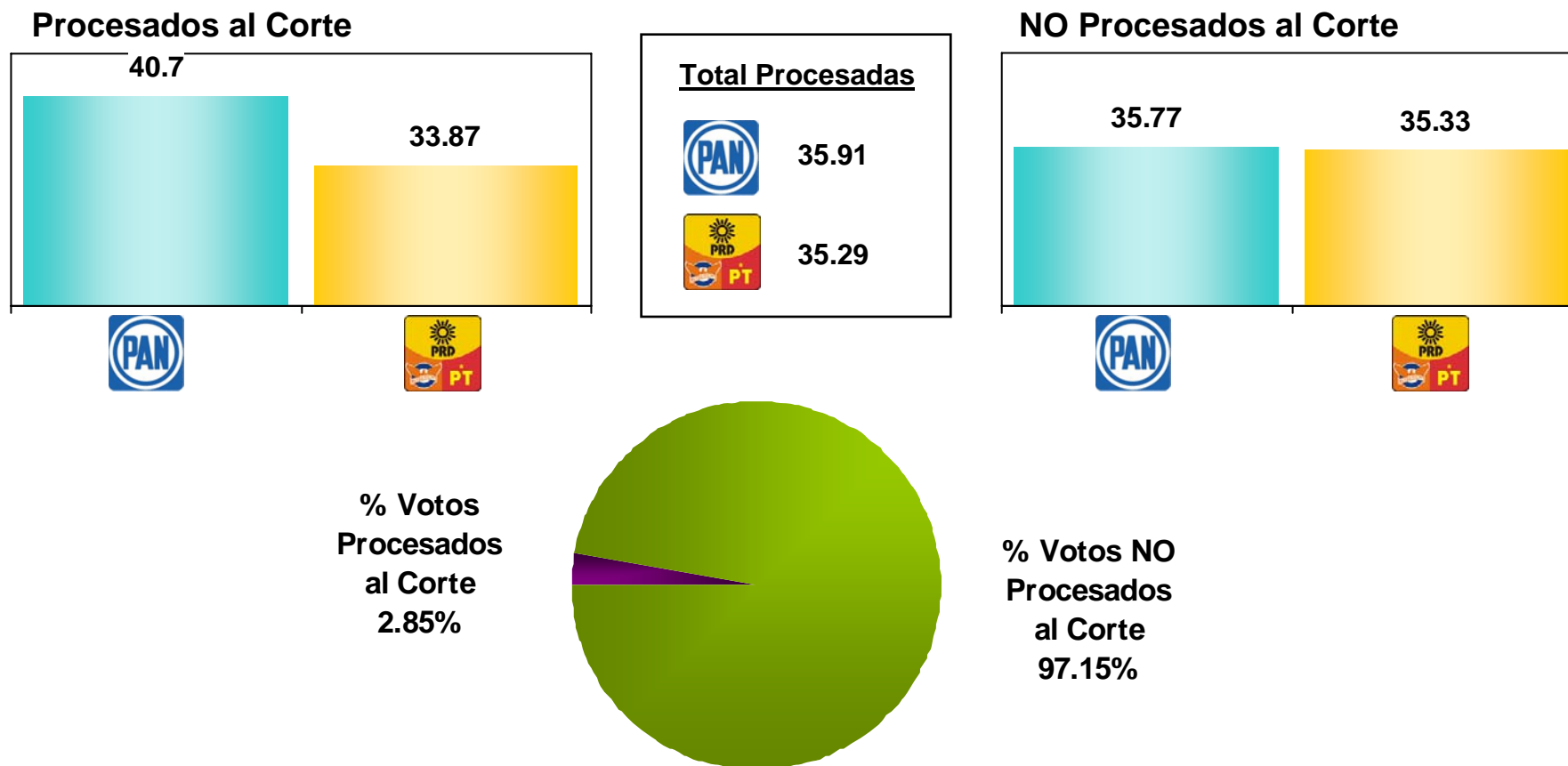
Corte 2. 21:00 hrs. del 2-julio-2006



Para el corte de las 21:00 hrs. el PAN alcanza su mayor nivel en las preferencias de los votos procesados

Corte 2. Acumulado a las 21:00 hrs del 2-julio-2006

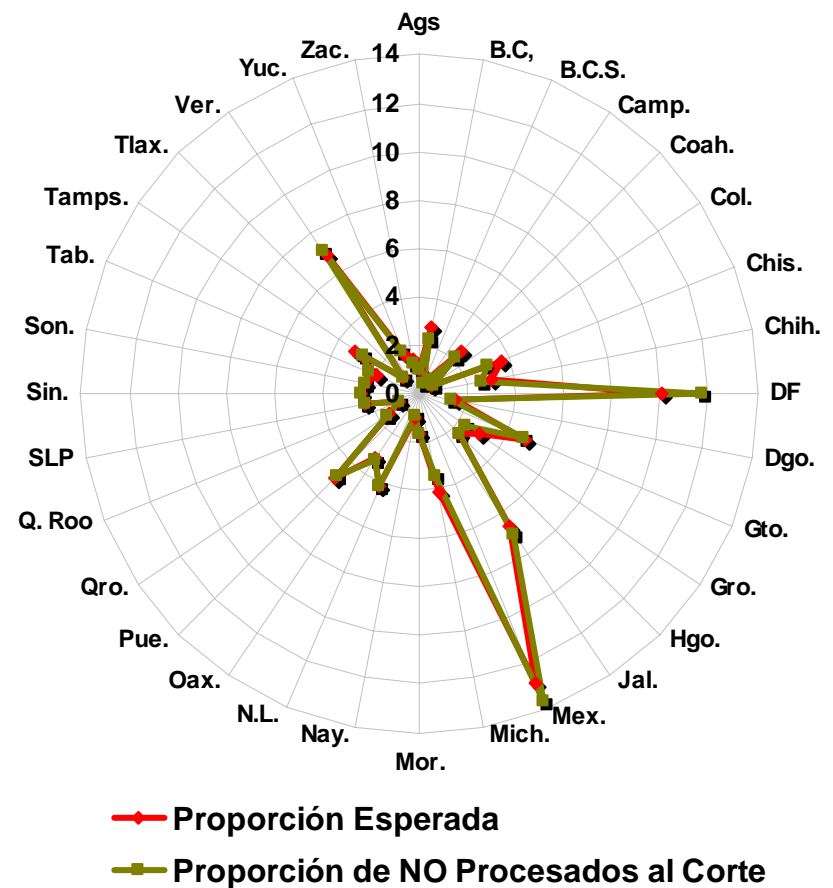
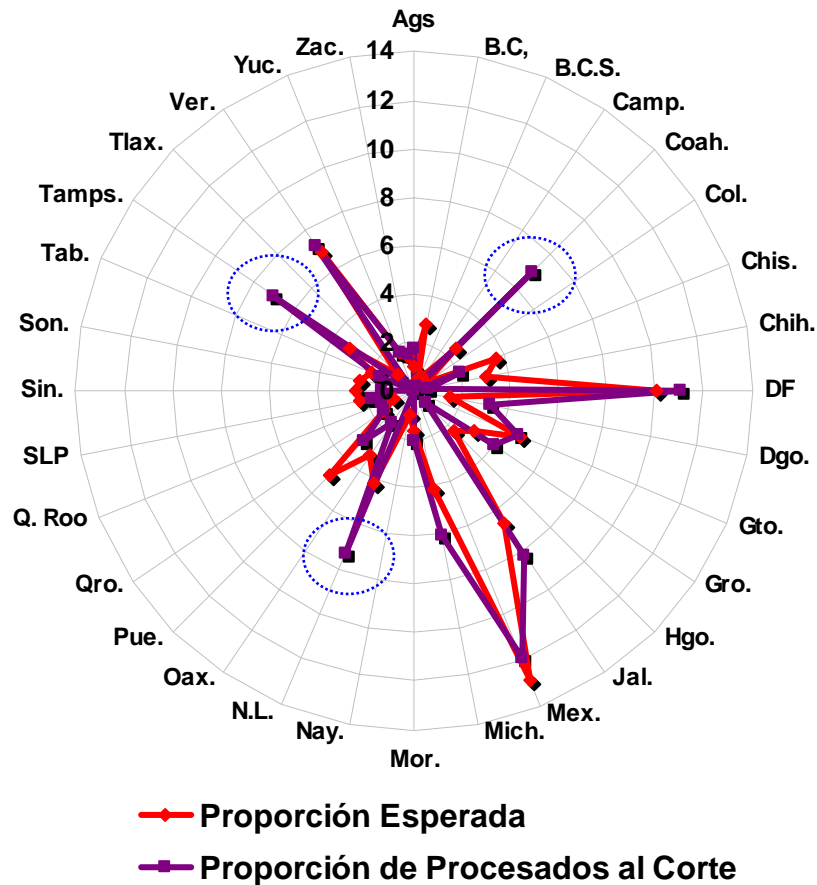
- Resultados de las Proporciones de Votos a Favor de los Dos Principales Candidatos



2 de los 3 estados con proporciones por arriba de las esperadas son Coahuila y Nuevo León (estados con preferencias Panistas)

Corte 2. Acumulado a las 21:00 hrs del 2-julio-2006

- % de Votos por Entidad Federativa con Respecto al Total Nacional de Votos

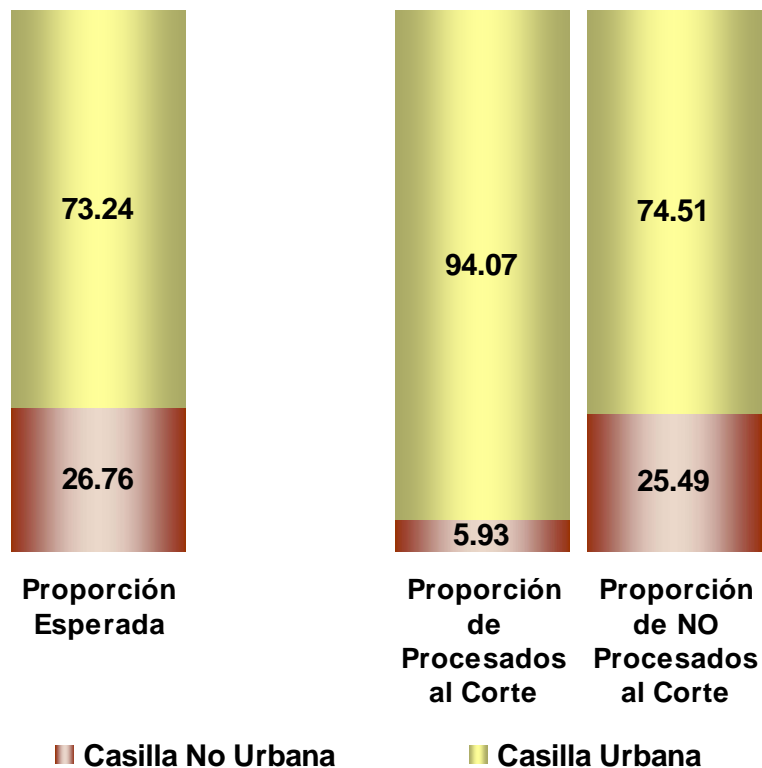


Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

La proporción de casillas Urbanas es mayor a lo esperado y las pruebas de bondad de ajuste mantienen que las proporciones esperadas, en las cifras procesadas, no ajustan

Corte 2. Acumulado a las 21:00 hrs del 2-julio-2006

- % de Votos por Tipo de Casilla con Respecto al Total Nacional de Votos



- Resultados Obtenidos por la Prueba de Bondad de Ajuste (ji-cuadrada)

Para las 32 entidades federativas

¿Las proporciones son iguales?		χ^2
Para los votos procesados al corte	NO	13.4 x10 ⁶
Para los votos NO procesados al corte	SÍ	1.116

Para los 2 tipos de casillas

¿Las proporciones son iguales?		χ^2
Para los votos procesados al corte	NO	77.776
Para los votos NO procesados al corte	SÍ	0.085



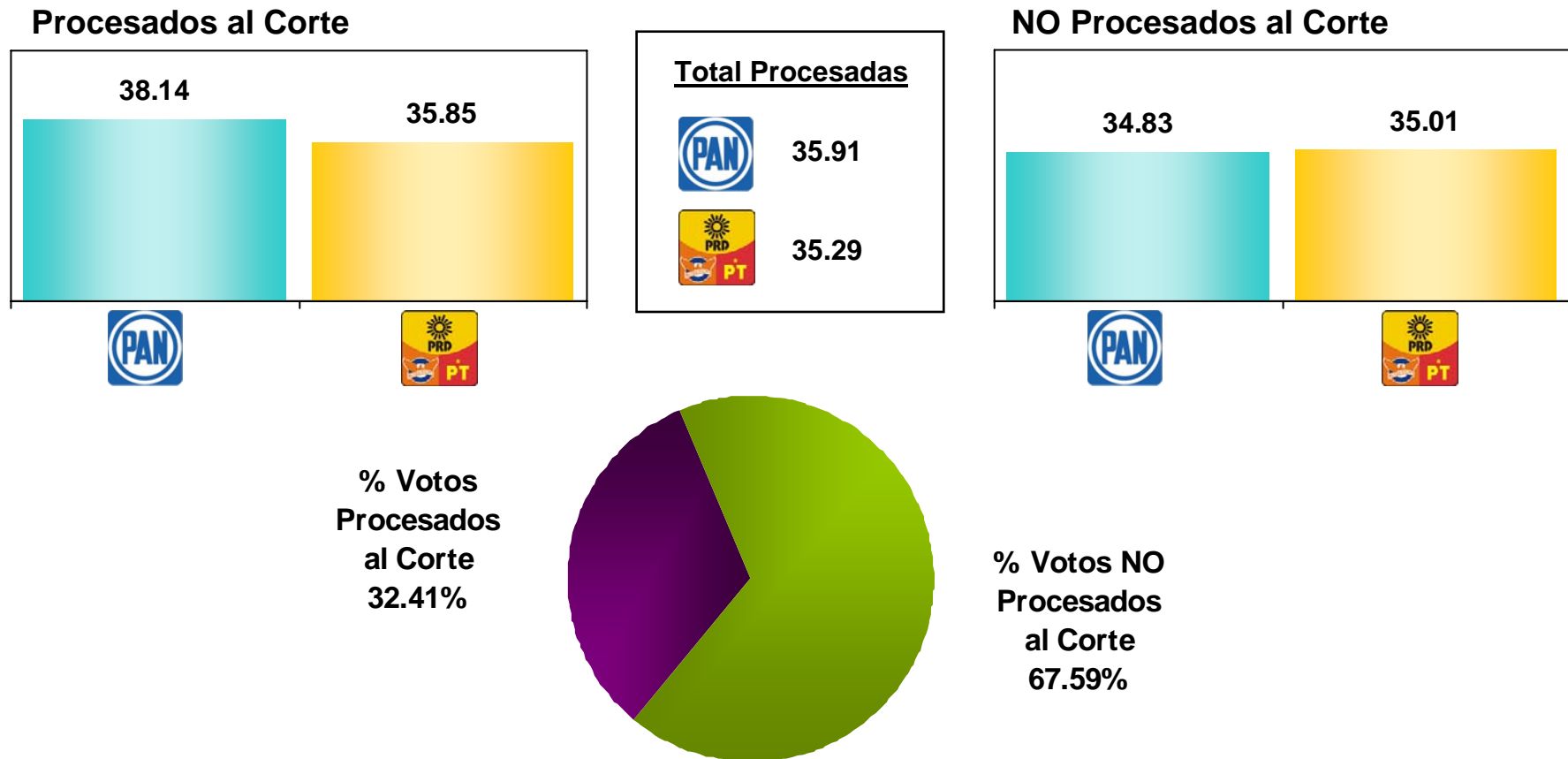
Corte 3. 23:00 hrs. del 2-julio-2006



A las 23:00 horas se ha computado cerca de una tercera parte de los votos y el PAN se mantiene con una ventaja sobre la coalición Por el Bien de Todos

Corte 3. Acumulado a las 23:00 hrs del 2-julio-2006

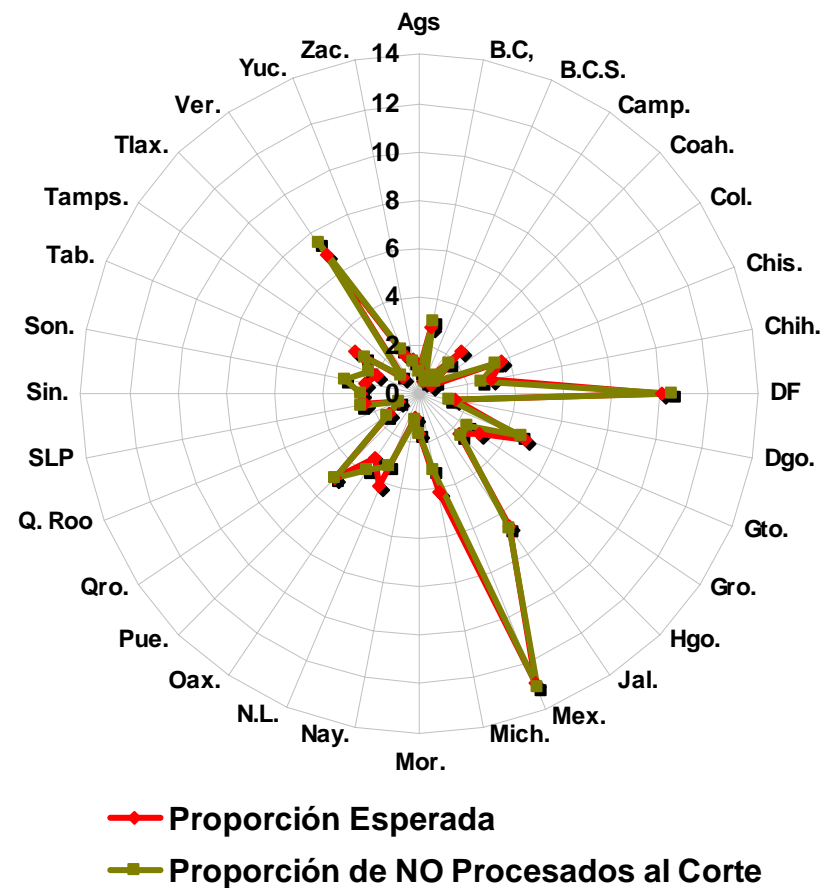
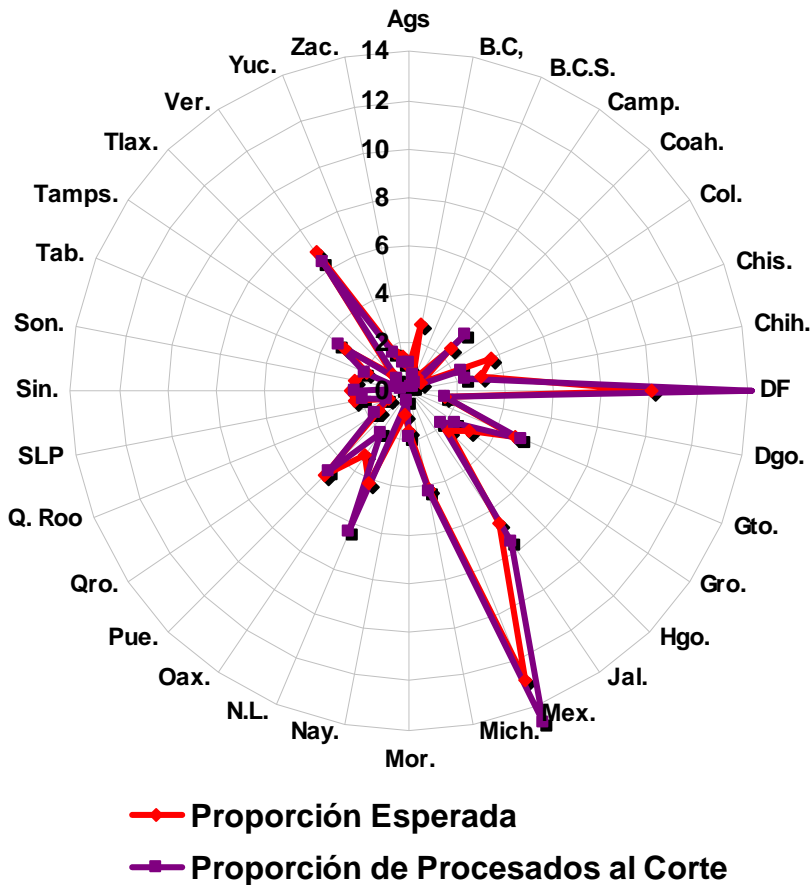
- Resultados de las Proporciones de Votos a Favor de los Dos Principales Candidatos



Las proporciones por entidad de los votos procesados se acerca a lo esperado, con excepciones como Coahuila, Tamaulipas o el DF

Corte 3. Acumulado a las 23:00 hrs del 2-julio-2006

- % de Votos por Entidad Federativa con Respecto al Total Nacional de Votos

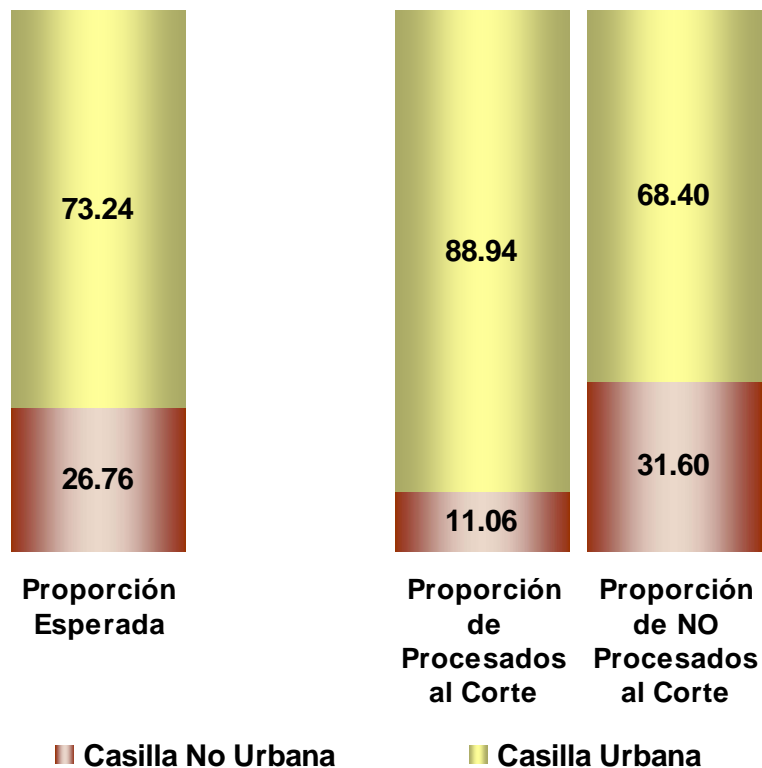


Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

En el tipo de casilla todavía se observan mayores proporciones que las esperadas para las casillas urbanas. Para este corte, la prueba corrobora el ajuste a nivel entidad

Corte 3. Acumulado a las 23:00 hrs del 2-julio-2006

- % de Votos por Tipo de Casilla con Respecto al Total Nacional de Votos



- Resultados Obtenidos por la Prueba de Bondad de Ajuste (ji-cuadrada)

Para las 32 entidades federativas

¿Las proporciones son iguales?		χ^2
Para los votos procesados al corte	SÍ	26.53
Para los votos NO procesados al corte	SÍ	1.817

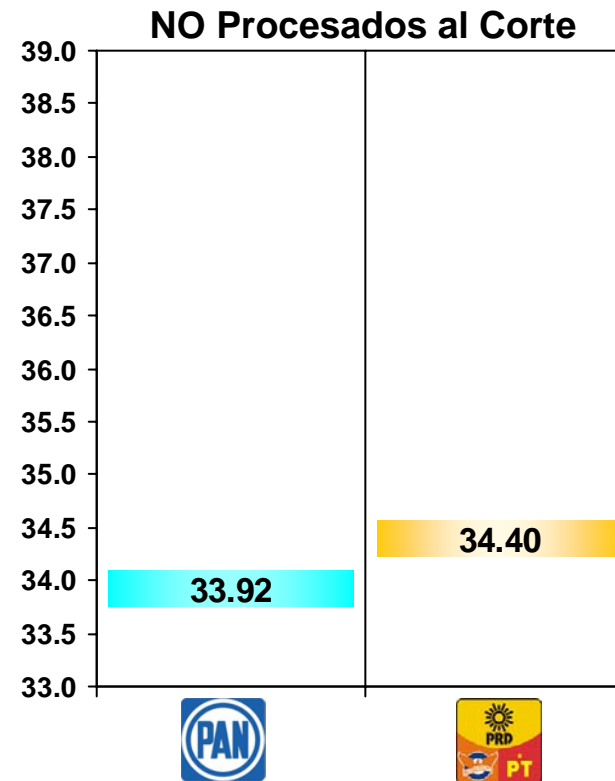
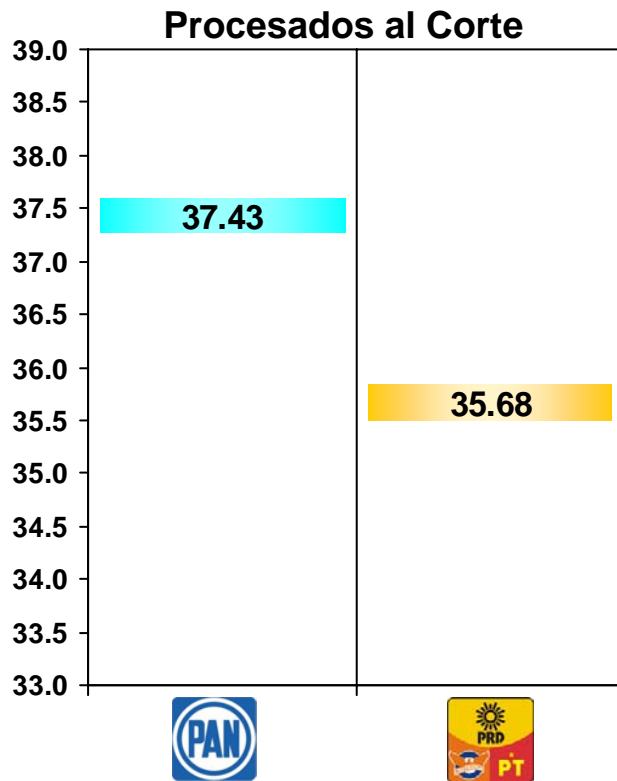
Para los 2 tipos de casillas

¿Las proporciones son iguales?		χ^2
Para los votos procesados al corte	NO	25.052
Para los votos NO procesados al corte	SÍ	1.081

Al analizar el nivel y la variabilidad de las proporciones obtenidas por cada candidato se observa que para las cifras aún no procesadas la distancia entre ambos es casi nula

Corte 3. Acumulado a las 23:00 hrs del 2-julio-2006

- Resultados de las Proporciones de Votos a Favor de los Dos Principales Candidatos y sus Intervalos de Confianza





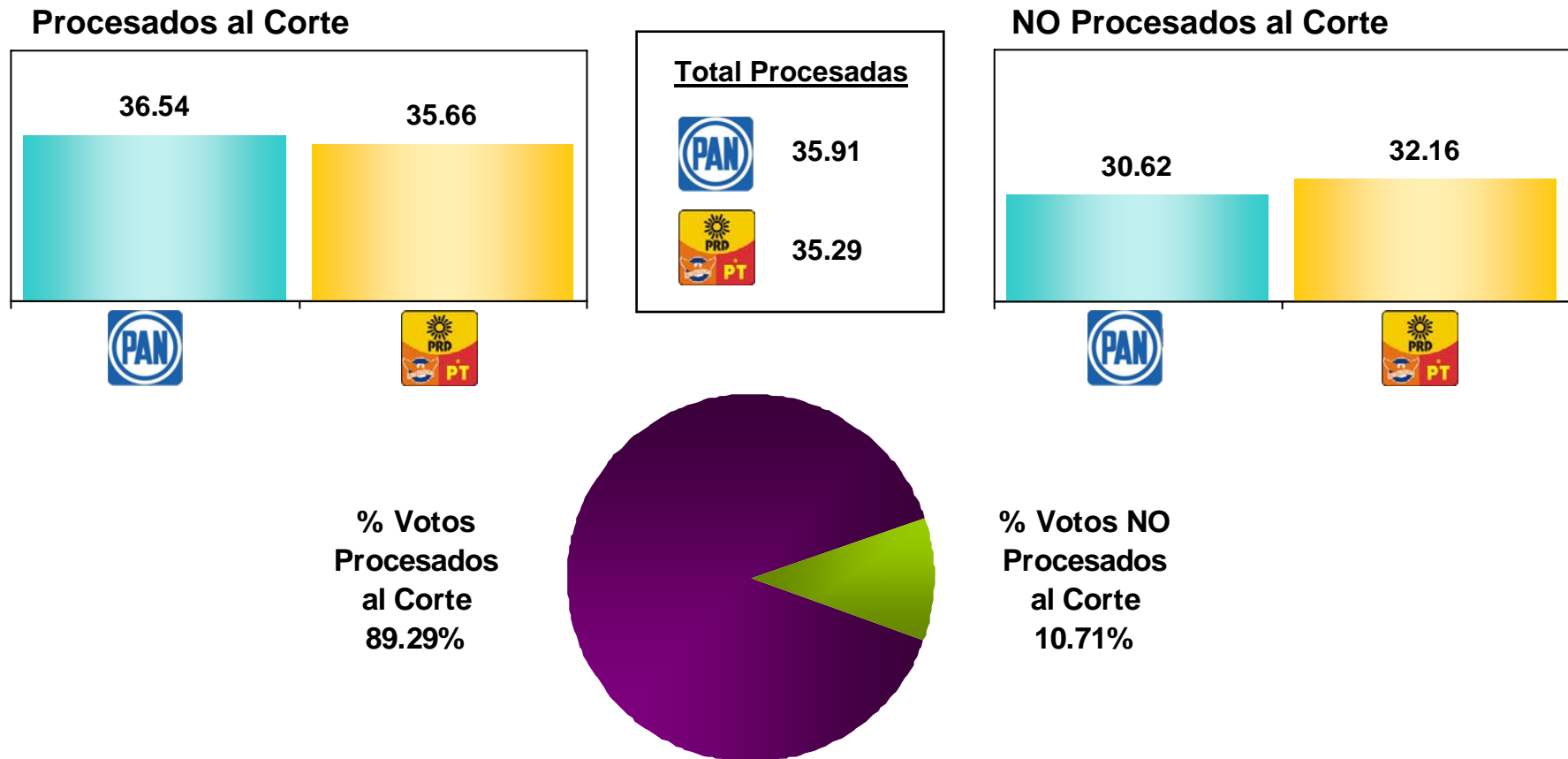
Corte 4. 04:40 hrs. del 3-julio-2006



A la 4:40 hrs del 3 de julio se ha procesado más del 80% de los votos y aunque la distancia entre los dos principales partidos es menor, el PAN se mantiene por arriba

Corte 4. Acumulado a las 04:40 hrs. del 3-julio-2006

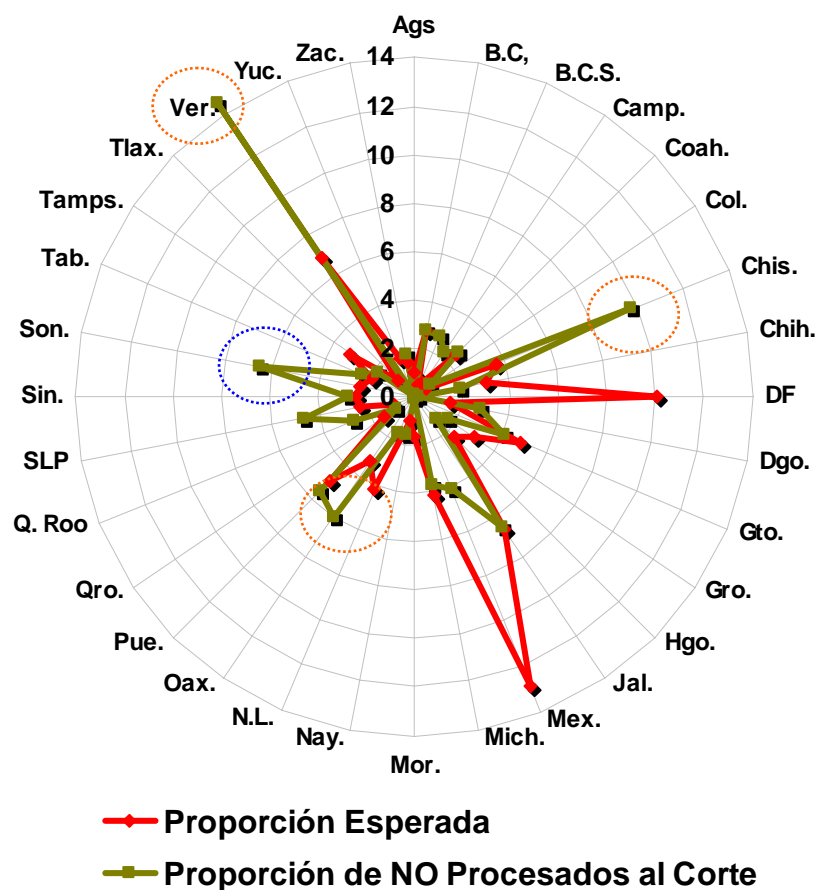
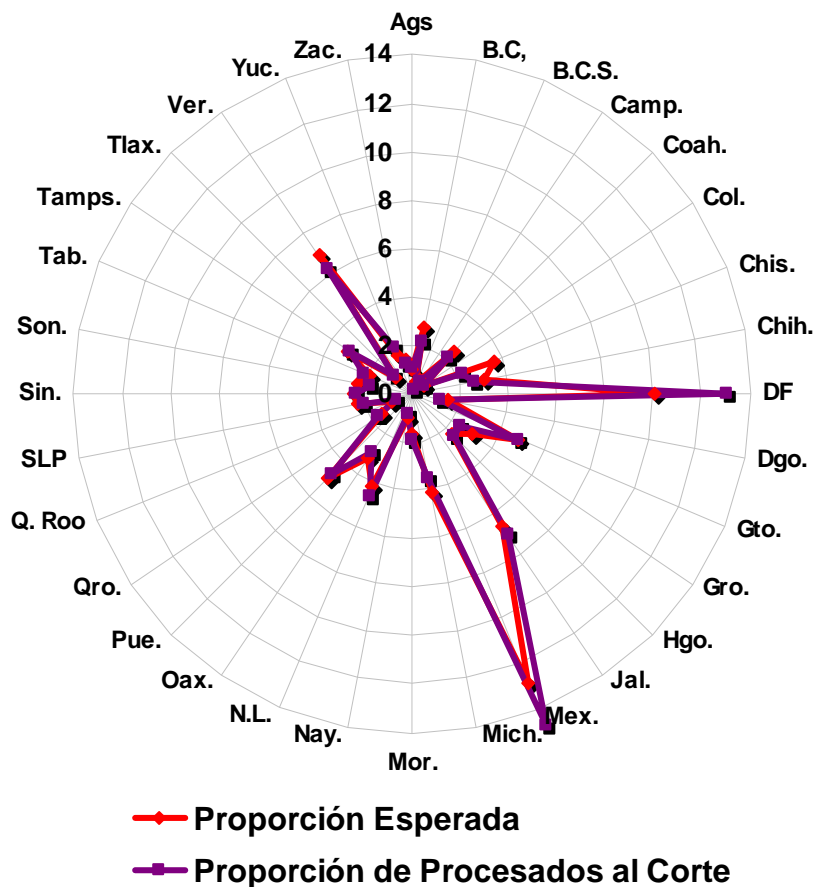
- Resultados de las Proporciones de Votos a Favor de los Dos Principales Candidatos



Las proporciones obtenidas de los votos procesados y las esperadas, son muy similares a nivel entidad

Corte 4. Acumulado a las 04:40 hrs. del 3-julio-2006

- % de Votos por Entidad Federativa con Respecto al Total Nacional de Votos

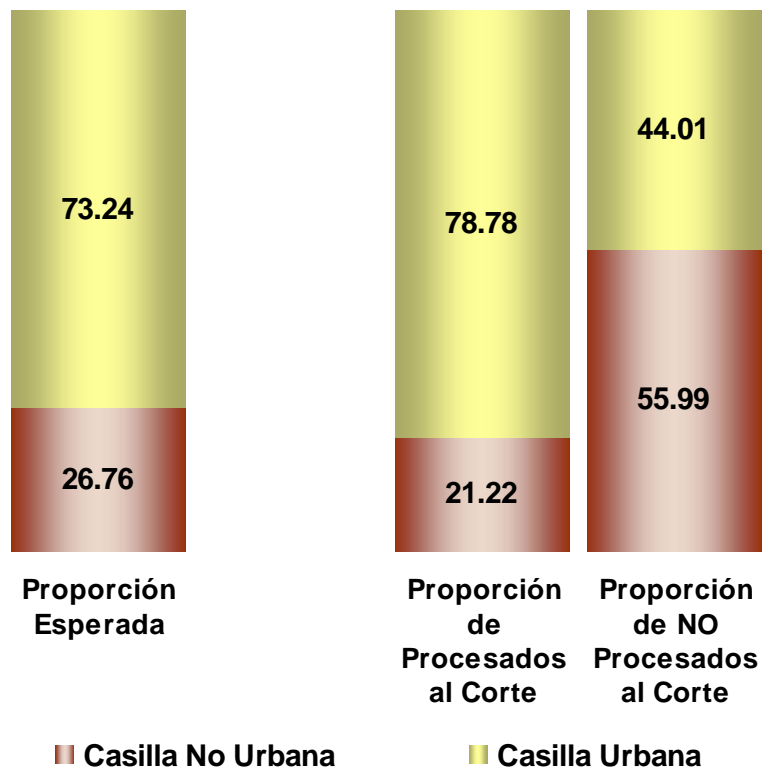


Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

Aunque se ha reducido, la proporción de casillas urbanas sigue estando por arriba de lo esperado, aunque ya no lo suficiente como para que la prueba rechace la igualdad

Corte 4. Acumulado a las 04:40 hrs. del 3-julio-2006

- % de Votos por Tipo de Casilla con Respecto al Total Nacional de Votos



- Resultados Obtenidos por la Prueba de Bondad de Ajuste (ji-cuadrada)

Para las 32 entidades federativas

¿Las proporciones son iguales?		χ^2
Para los votos procesados al corte	SÍ	3.800
Para los votos NO procesados al corte	NO	9.99×10^5

Para los 2 tipos de casillas

¿Las proporciones son iguales?		χ^2
Para los votos procesados al corte	SÍ	1.839
Para los votos NO procesados al corte	NO	34.652



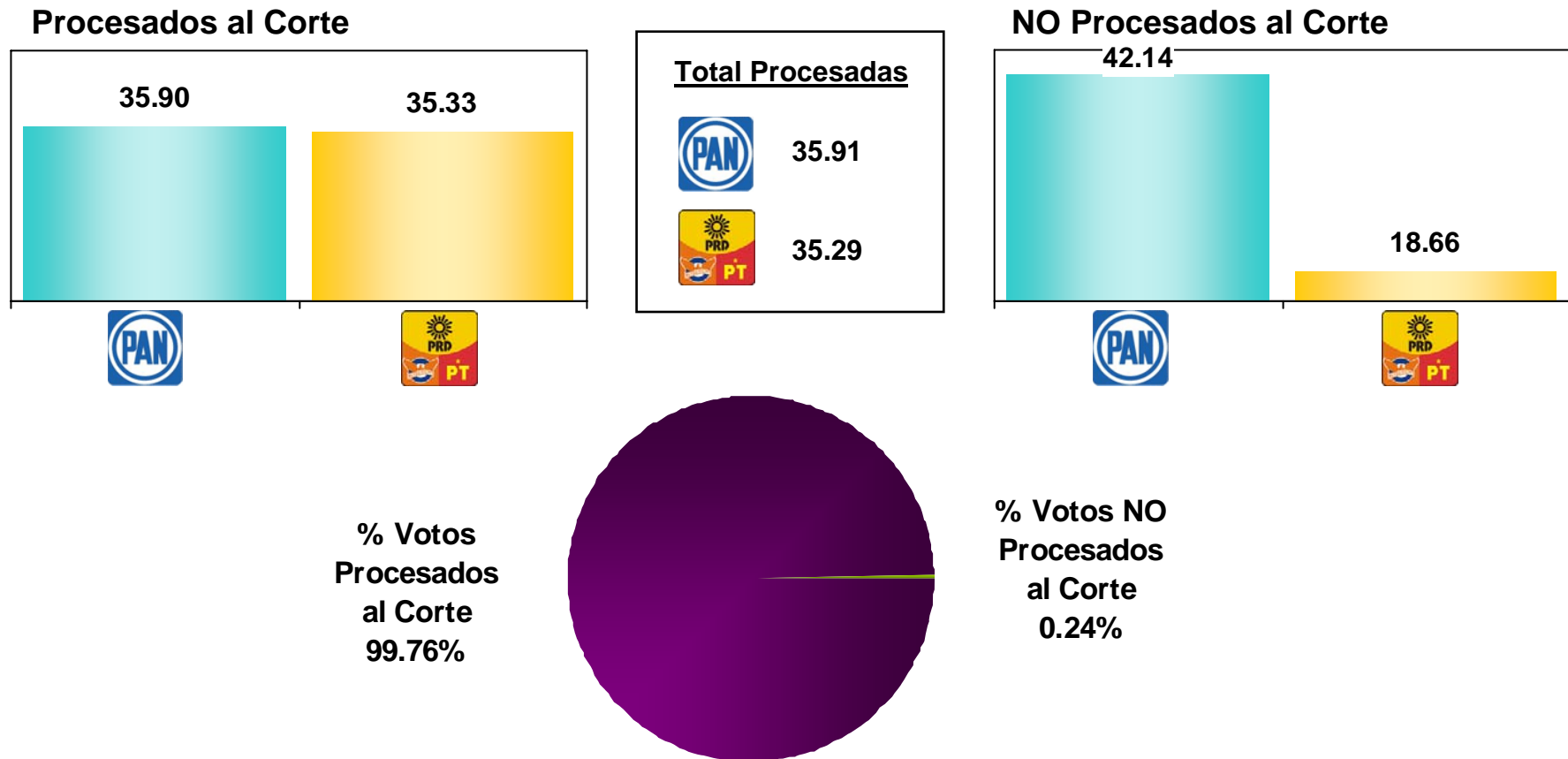
Corte 5. 15:40 hrs. del 3-julio-2006



A las 15:40 hrs. prácticamente se han procesado el 100% de los votos del 94.5% de las actas y se observa que en los votos pendientes se da una mayor preferencia por el PAN

Corte 5. Acumulado a las 15:40 hrs. del 3-julio-2006

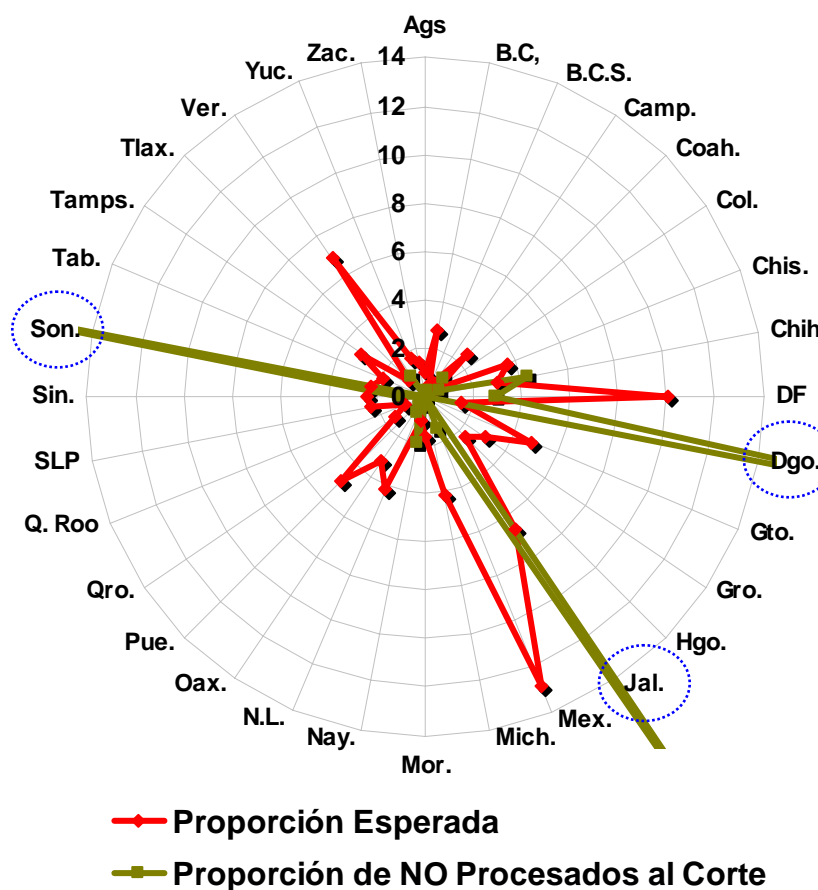
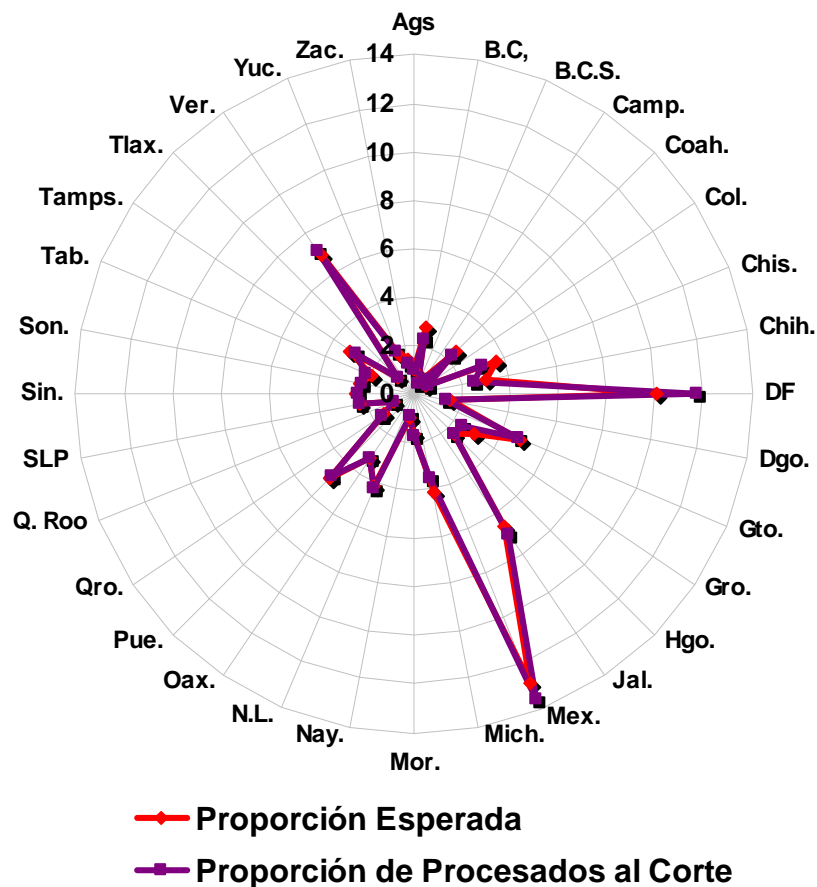
- Resultados de las Proporciones de Votos a Favor de los Dos Principales Candidatos



Analizando las entidades con proporciones altas, en relación a lo esperado, de los votos pendientes de procesar son Sonora, Durango y Jalisco

Corte 5. Acumulado a las 15:40 hrs. del 3-julio-2006

- % de Votos por Entidad Federativa con Respecto al Total Nacional de Votos

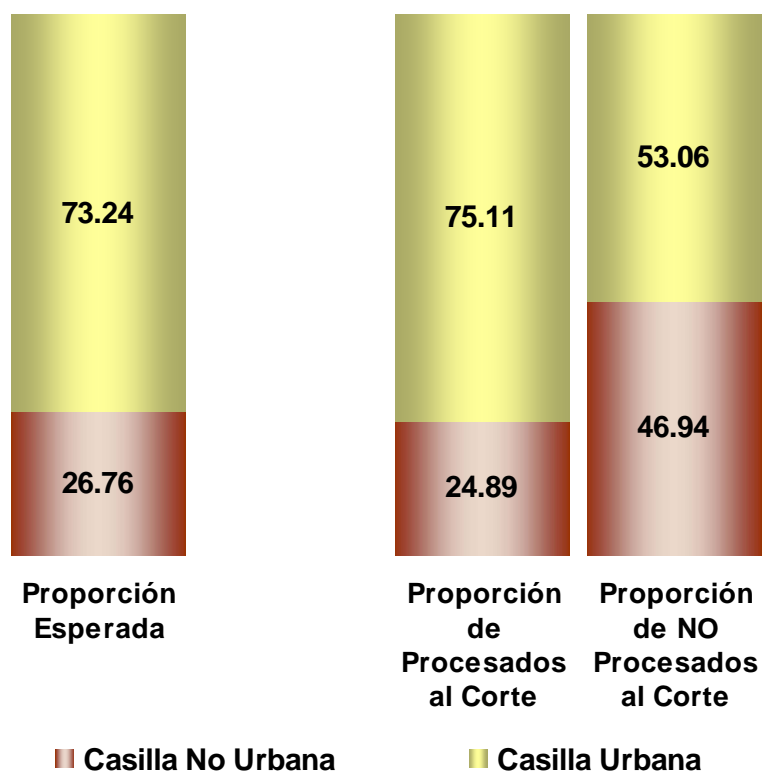


Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

Y al correr la prueba para la proporción pendiente de procesar se tiene que las proporciones (por tipo de casilla y entidad) no concuerdan con lo esperado

Corte 5. Acumulado a las 15:40 hrs. del 3-julio-2006

- % de Votos por Tipo de Casilla con Respecto al Total Nacional de Votos



- Resultados Obtenidos por la Prueba de Bondad de Ajuste (ji-cuadrada)

Para las 32 entidades federativas

¿Las proporciones son iguales?		χ^2
Para los votos procesados al corte	SÍ	1.072
Para los votos NO procesados al corte	NO	10.16 x10 ⁷

Para los 2 tipos de casillas

¿Las proporciones son iguales?		χ^2
Para los votos procesados al corte	SÍ	0.187
Para los votos NO procesados al corte	NO	16.34

Sumario

- El origen de resultados en la votación que se antojan atípicos, con respecto al resultado final, a distintos cortes en el tiempo, está correlacionado con la representatividad que pudieran tener las cifras procesadas hasta ese corte
- Al realizar pruebas de bondad de ajuste en 5 diferentes cortes, se encontró que en aquellos cortes en los que las cifras procesadas presentaban variaciones más altas, en relación a las finales, las pruebas rechazaban el ajuste, tanto a nivel entidad como a nivel tipo de casilla
- Para los resultados que se mostraban más cercanos a los procesados, se verificó que al menos el ajuste a nivel entidad se cumplía y que a medida que se cumplía el ajuste a nivel tipo de casilla el resultado era todavía más cercano al final

Sumario

- ¿Cuál sería una causa probable para observar una ventaja de López Obrador hasta antes de las 19:40 hrs. que no se volvió a observar?
 - De acuerdo al análisis del presente documento, el único momento en el que la coalición por el Bien de Todos mantenía una ventaja por encima del PAN, era en un corte en el que las proporciones observadas para los votos procesados no ajustaban con las proporciones esperadas
 - La prueba de bondad de ajuste mostraba que no existía evidencia que indicara que las proporciones observadas eran similares a las esperadas, lo que permitiría pensar que los datos con los que se contaba hasta ese momento difícilmente serían representativos del total

Sumario

- ¿A qué se le podría atribuir el hecho de que la ventaja de Calderón vs. López Obrador se reduzca pero que no se vuelva a dar un cruce entre las proporciones?
 - En el corte de las 21:00 hrs. del 2 de julio se observó el mejor nivel en las preferencias por parte del PAN
 - Al realizar pruebas de bondad sobre estas cifras se encontró que las proporciones eran distintas a las esperadas, por lo que se esperaba que las cifras con las que se contaba difícilmente eran representativas del resultado final
 - Al corte de las 23:00 hrs. del 2 de julio las pruebas de bondad de ajuste corroboran que se ha dado el ajuste en las proporciones esperadas, a nivel entidad para las cifras procesadas, así como para las pendientes de procesar
 - A este corte las cifras procesadas marcaban una ventaja de 2.29% a favor del PAN, mientras que las pendientes de procesar marcaban una ventaja para la coalición Por el Bien de Todos de 0.18%

Sumario

- Por la magnitud de las diferencias entre ambos grupos (procesados y pendientes de procesar) y el peso de cada uno de ellos, 32.4% y 67.6%, respectivamente, se tiene que la ventaja del PAN tenderá a reducirse pero no al nivel de quedar por debajo de la preferencia obtenida por la coalición Por el Bien de Todos

[Regresar a Reflexiones](#)

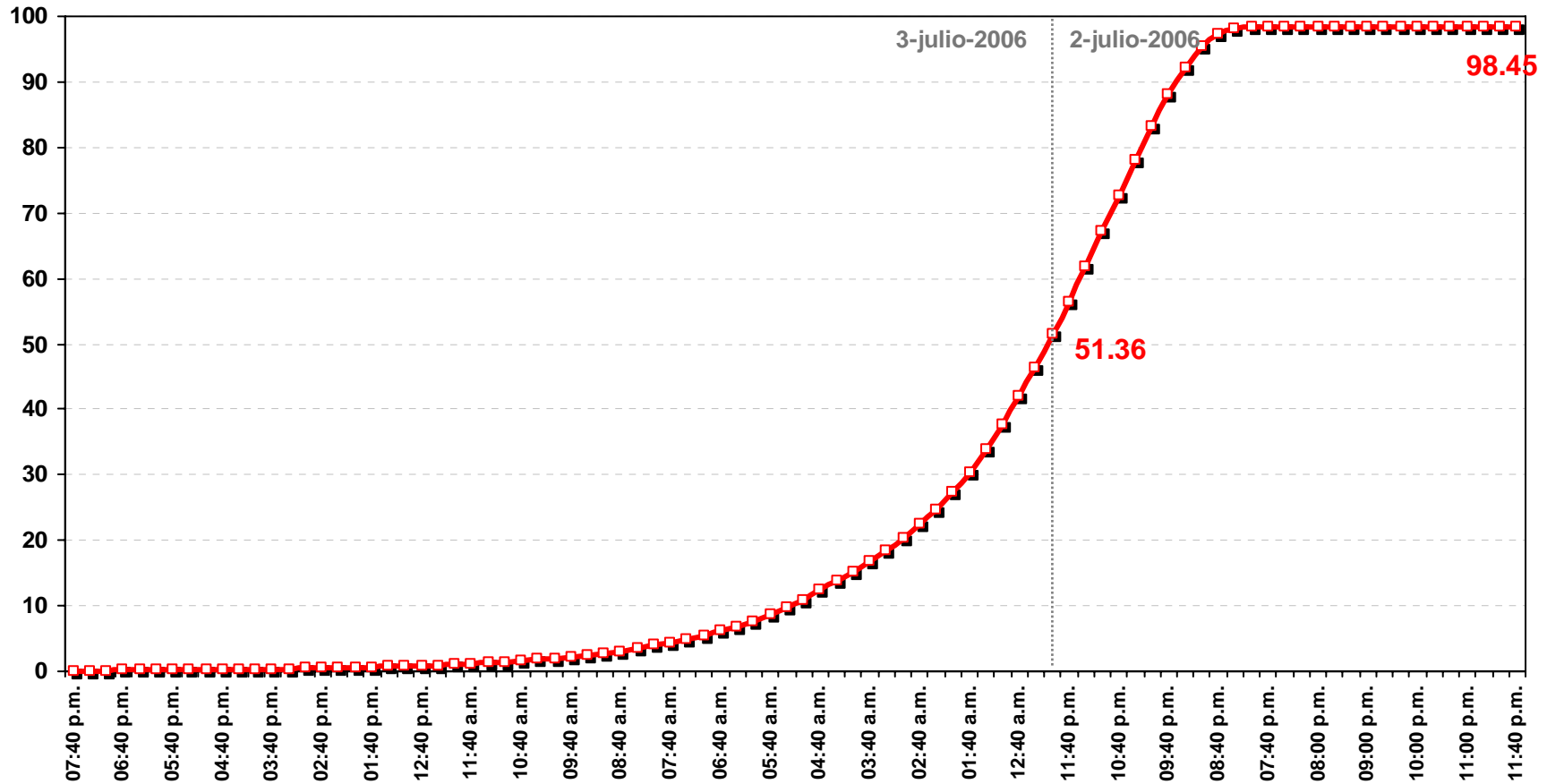


**¿Más cruces entre las preferencias indican
mayor certidumbre sobre las cifras?
Simulación de captura inversa**



Simulando una captura del PREP en sentido inverso al que llegaron las actas, se observa la baja contribución de las “primeras horas”

- % Acumulado de Actas Procesadas en Simulación Inversa por Hora de Registro en el PREP (Cortes de 20 min.)



Fuente: IFE

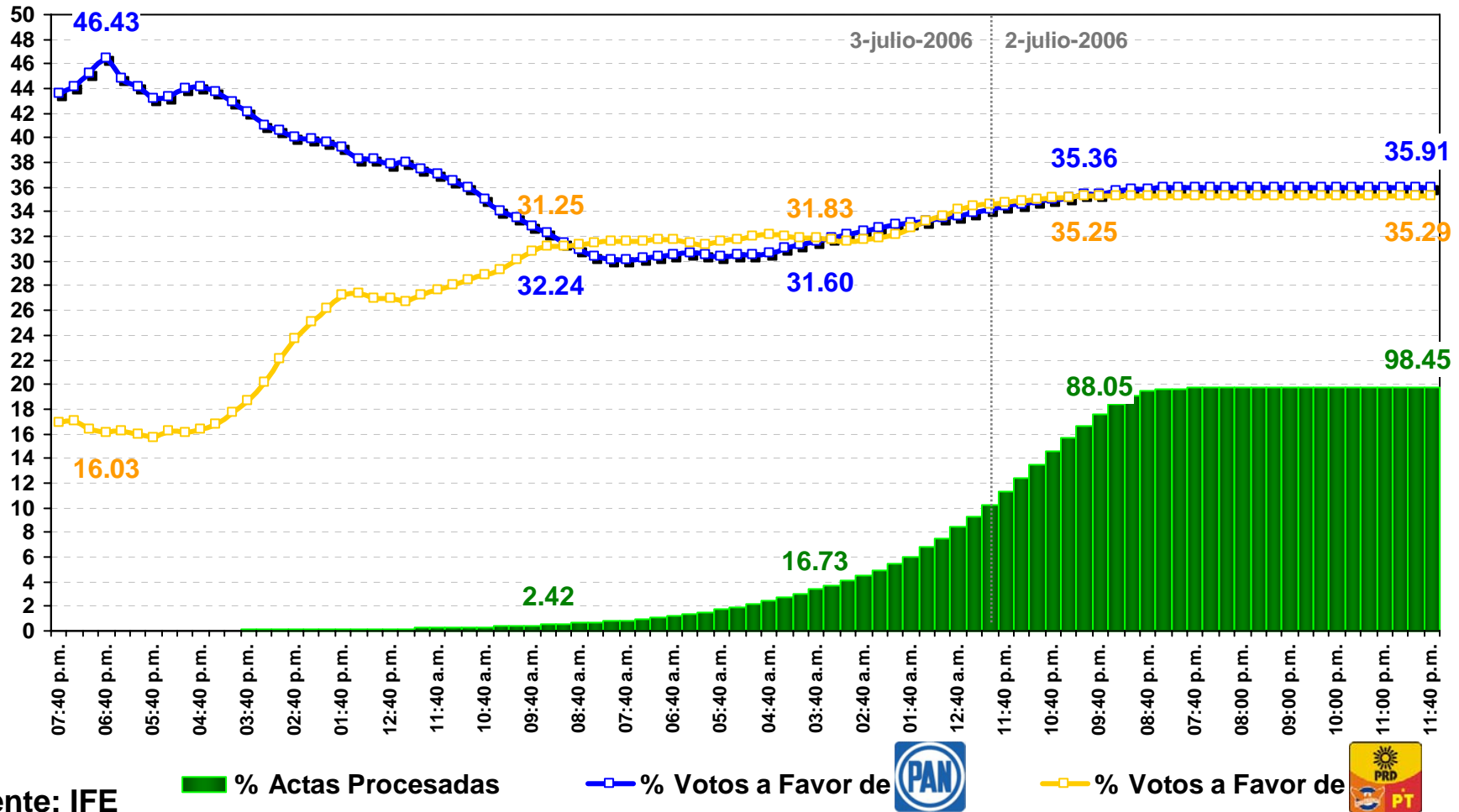
Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business



Elección 2006

En esta simulación se encuentra una mayor cantidad de cruces entre las proporciones obtenidas por el PAN y la coalición Por el Bien de Todos

- % Votos Obtenidos a Favor de los Dos Principales Candidatos a la Presidencia del País



Fuente: IFE

% Actas Procesadas

% Votos a Favor de

% Votos a Favor de

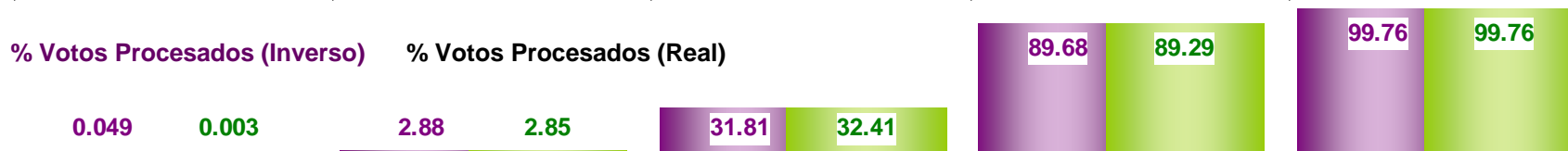
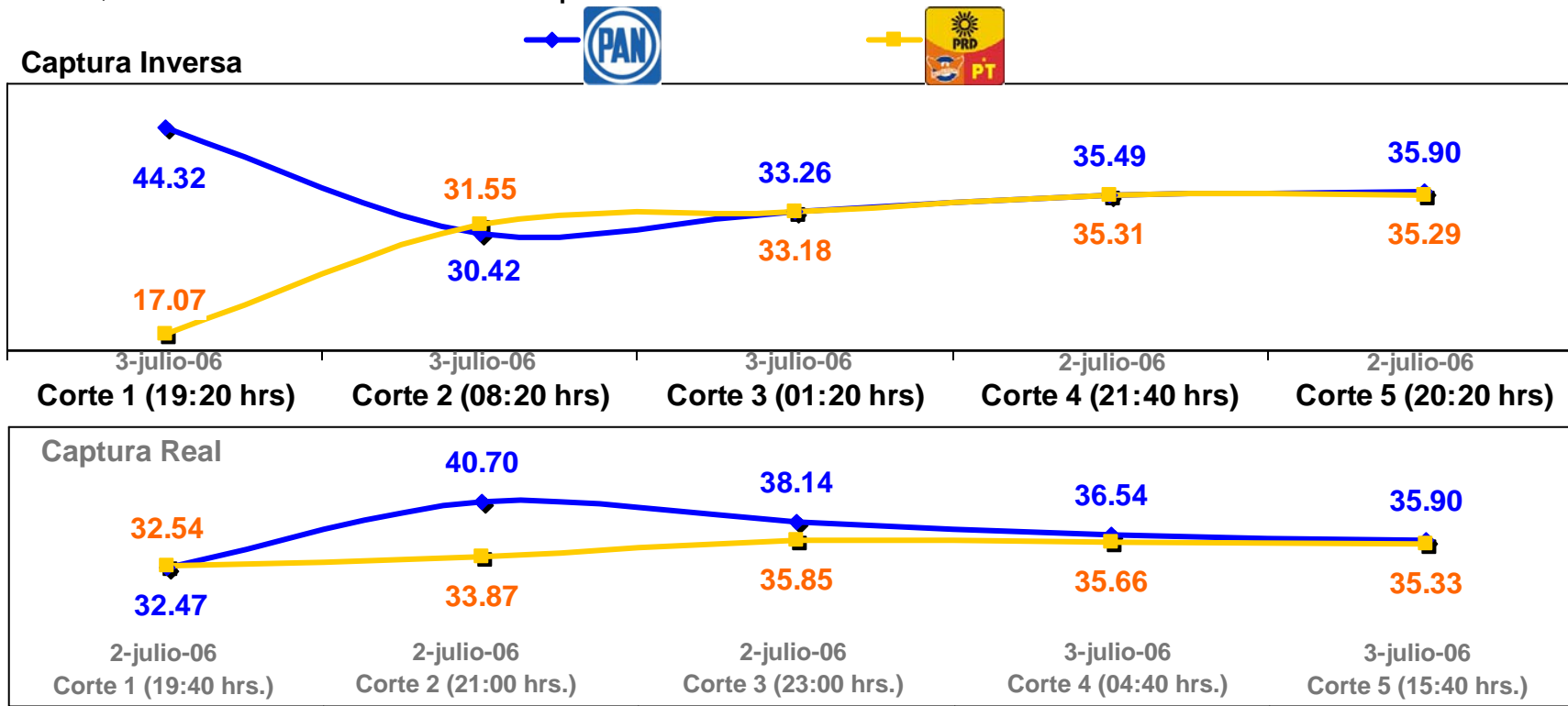
Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business



Elección 2006

A cortes similares, en cuanto a votos procesados, se habría observado una contienda más cerrada, pero evidentemente el resultado habría sido el mismo

- Resultados de las Proporciones Obtenidas a partir de los Votos Procesados, de Manera Inversa, a Favor de los Dos Principales Candidatos a cada Corte



Confidential & Proprietary • Copyright © 2006 ACNielsen • a VNU business

Sumario

- Suponiendo que las últimas actas procesadas por el PREP hubieran sido las primeras y manteniendo este patrón inverso, se hubieran observado cuatro cruces entre las preferencias de los dos principales partidos
- El último de estos cruces, se habría observado hasta poco antes de alcanzar el 80% de los actas procesadas. A partir de ese momento no se habrían observado más
- De lo anterior se tiene que:
 - Evidentemente, los resultados finales no varían por la existencia de una mayor cantidad de cruces
 - Las mismas actas procesadas de manera inversa, sin necesidad de un acomodo especial, generan una mayor cantidad de cruces, dejando ver que el cruce o no de las preferencias es un mero evento casuístico que va más relacionado con la velocidad con la que cada entidad y/o distrito enviaba sus cifras

[Regresar a Reflexiones](#)



¿La ley de Benford se puede aplicar a los votos en las casillas?



Antecedentes

- En 1881, el astrónomo y matemático Simon Newcomb observó que las primeras páginas de las tablas de logaritmos estaban manifiestamente más usadas que las finales
 - De lo anterior, dedujo que aparentemente los dígitos iniciales de los números no son equiprobables, sino que el 1 aparece como dígito inicial más frecuente, seguido del 2, etc. hasta que el 9 es el menos frecuente
- En 1938, de manera independiente, el físico Frank Benford observó el mismo fenómeno y realizó una comprobación empírica sobre un total de 20,229 números agrupados en 20 muestras de gran diversidad y a partir de estos resultados postuló una “ley de los números anómalos”

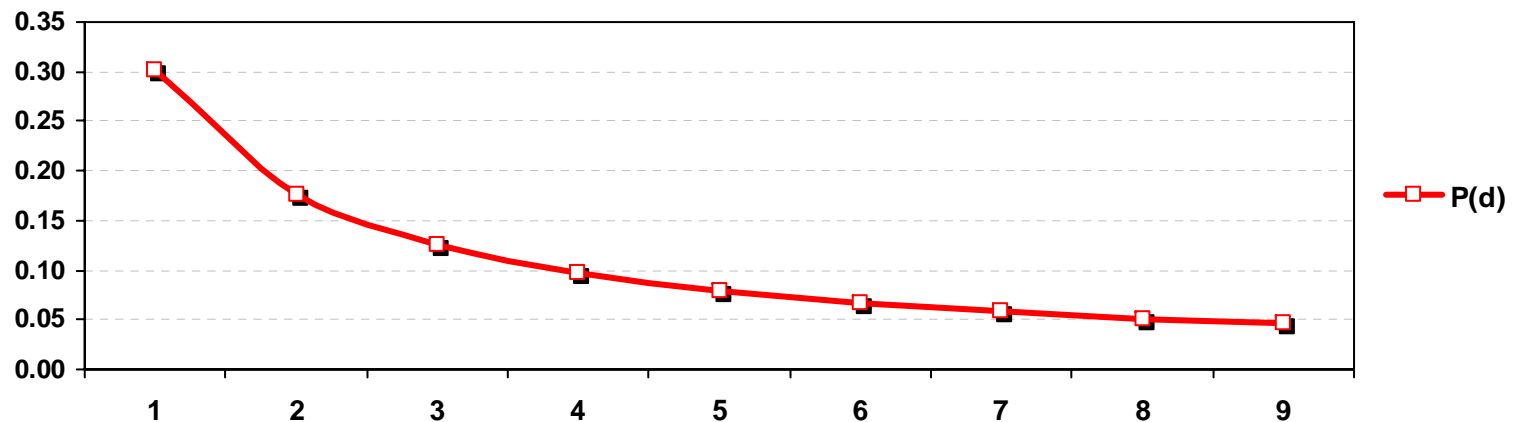
Ley de Benford

- La “ley de Benford” es una ley logarítmica que puede enunciarse como:

$$P(d) = \log\left(\frac{d+1}{d}\right) = \log(d+1) - \log(d) \quad d \in \{1, 2, \dots, 9\}$$

- De aquí se deriva que las probabilidades de los primeros dígitos sea la siguiente:

Dígito d	0	1	2	3	4	5	6	7	8	9
$P(d)$		0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046



¿Qué datos satisfacen la ley de Benford?

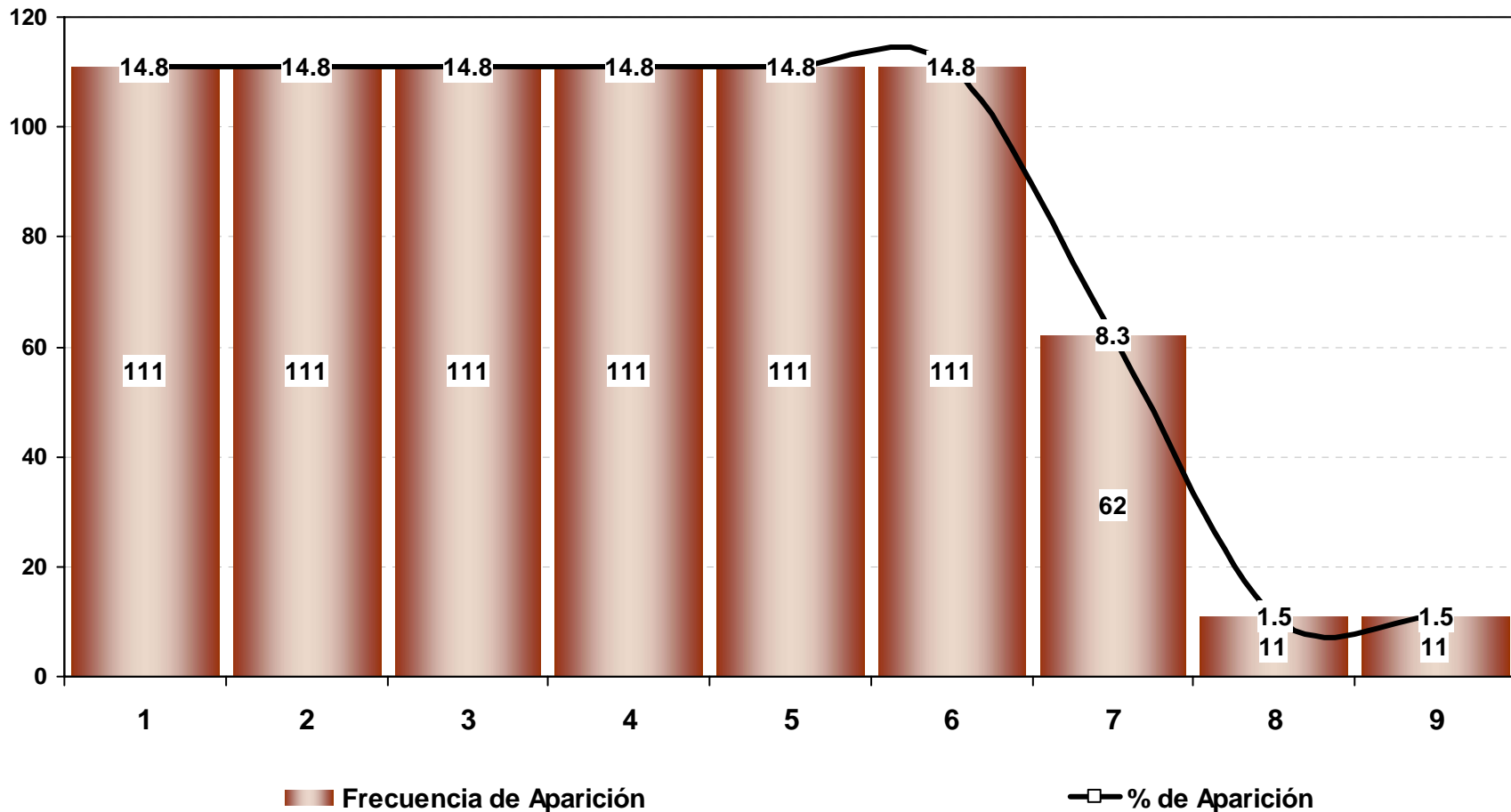
- De acuerdo a diversos autores, es evidente que la ley de Benford no se verifica en todos los posibles conjuntos de datos numéricos:
 - Aquellos procedentes de distribuciones uniformes (*i.e.* números de lotería) o normales (*i.e.* edades de personas)
 - Cuando los datos tienen limitado el valor del dígito inicial
- Considerando exclusivamente datos de origen matemático se ha comprobado que los números procedentes de evaluar funciones comunes como x^2 , $x^{1/2}$ ó $1/x$ no verifican la ley
- En pocas palabras, existe una fuerte dependencia en cuanto a la naturaleza de los datos
 - Es seguro que números como los telefónicos o los de documentos de identidad no siguen la distribución logarítmica, pues se asignan arbitrariamente

Aplicación de la ley de Benford

- Una de las aplicaciones más comunes de la ley de Benford se da en auditorías contables, para detectar fraudes
- Pero incluso en estos casos algunos autores coinciden en que se debe tener en cuenta ciertas características de las cifras para que la ley de Benford aplique:
 - Las cifras deben medir el mismo atributo
 - No debe haber un valor mínimo o máximo preestablecido
 - Las cifras deben ser aleatorias y no preasignadas
 - No debe existir una aglomeración de cifras alrededor de un cierto valor
- Con base en lo anterior ¿la ley de Benford puede ser aplicada para los resultados electorales de la elección presidencial del 2006?

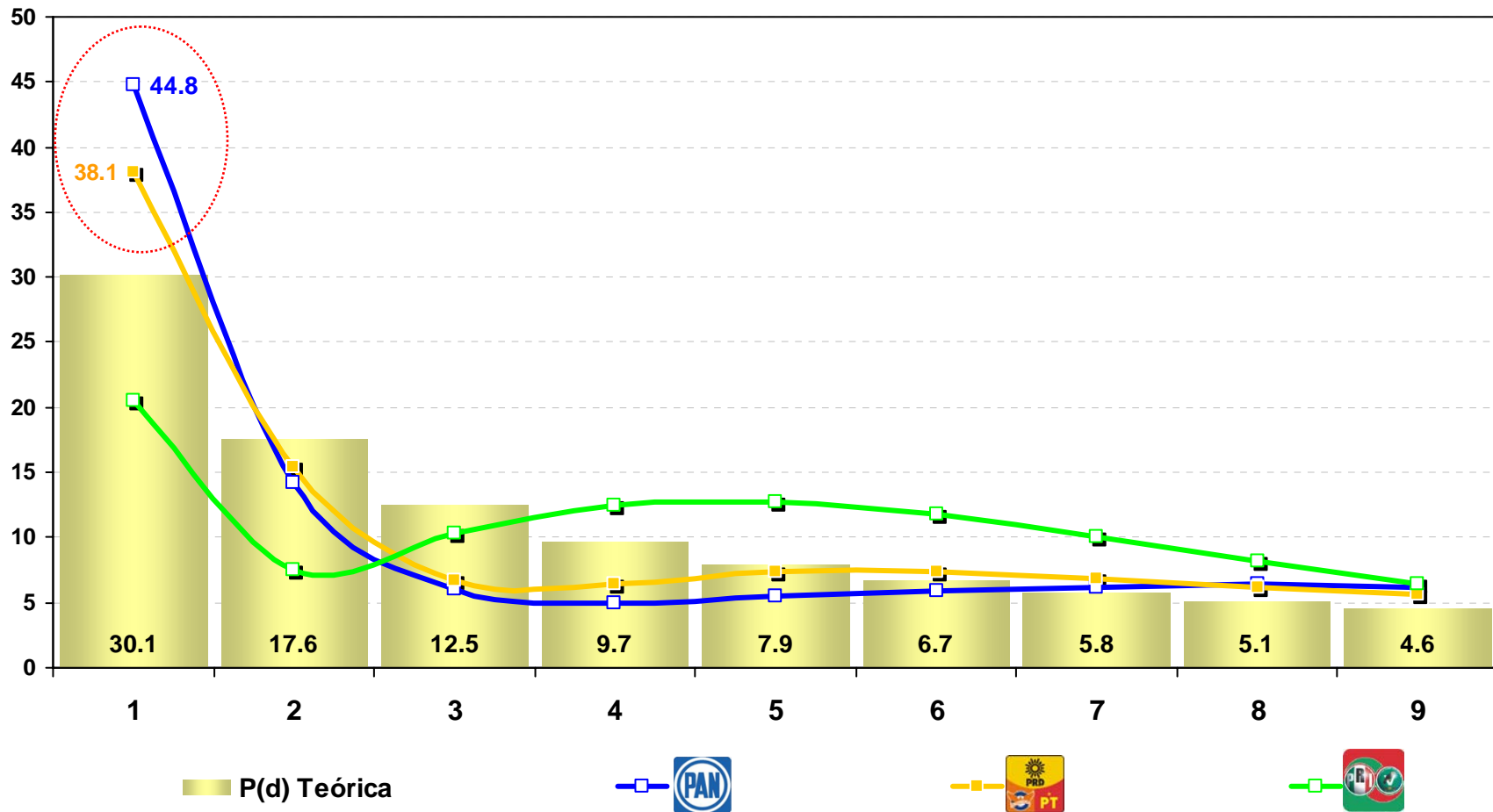
La cantidad máxima de boletas por casilla era de 750, lo que acota la aparición del 7, el 8 y el 9 como primeros dígitos, restringiendo el uso de la ley de Benford

- Frecuencia de aparición como primer dígito, para cada uno de los dígitos, en una serie que va del 1 al 750



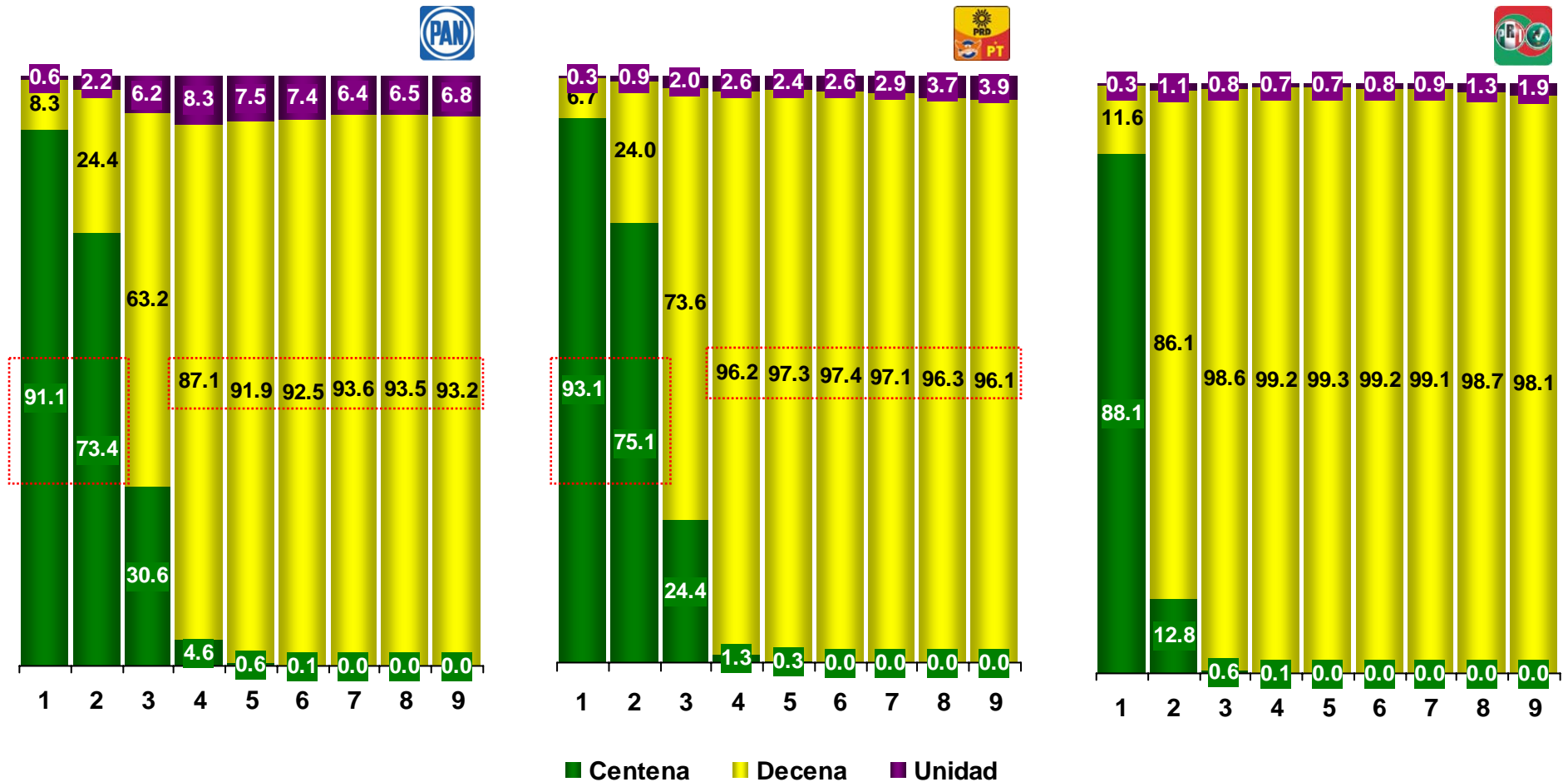
Y observando las proporciones del primer dígito de los votos por casilla, para cada partido, se observa que efectivamente la ley de Benford no se cumple

- % de apariciones de cada dígito, como primer dígito, en los votos a partir de cada una de las casillas procesadas por el PREP (128,771 casillas)



Cabe mencionar que la coalición Por el Bien de Todos obtuvo una composición de centenas y decenas, relativamente mayor a la obtenida por el PAN con excepción del dígito 3

- Proporción de las posiciones que ocupan los primeros dígitos de cada partido



Sumario

- Diversos autores sostienen que la ley de Benford no se verifica en todos los posibles conjuntos numéricos
 - Una de las restricciones que se sostienen es que los valores del conjunto no deben estar acotados
- En las elecciones presidenciales del 2006, el número máximo de boletas por casilla era de 750, acotando la aparición del 7,8 y 9.
 - Lo anterior implica que la ley de Benford no aplica necesariamente a este conjunto de datos
- Al analizar las cifras, se comprueba que efectivamente la ley de Benford no se cumple para los primeros dígitos, de los votos por casilla, obtenidos por los tres principales partidos
 - Adicionalmente, se observa que en algunos dígitos en los que existe mayor discrepancia con las frecuencias propuestas por la ley, la coalición Por el Bien de Todos alcanza proporciones mayores de decenas y centenas que cualquiera de los otros dos partidos

[Regresar a Reflexiones](#)



AC Nielsen
Conclusiones



Conclusiones

- En diversos ejercicios muestrales se encontró que ante la cercanía de los resultados obtenidos por el PAN y la coalición Por el Bien de Todos, existían muestras que favorecían a uno u otro partido
 - Sin embargo, la proporción de muestras a favor de la coalición Por el Bien de Todos fue menor que la proporción de muestras a favor del PAN
 - Asimismo, se encontró que ante la dispersión que se tenía alrededor de las medias, en estos casos no era posible determinar que una estimación fuera superior a la otra, es decir, que no se podía determinar a un partido vencedor
- Dada la naturaleza operativa del PREP, tomar cortes aleatorios en distintos momentos, no asegura contar con una muestra representativa del comportamiento nacional
 - De acuerdo a los análisis que se realizaron, si fuera necesario requerir de una muestra, esta sólo podría obtenerse a partir de las 4:40 a.m. del 3 de julio de 2006, cuando ya se hubiera procesado poco más del 85% de las actas

Conclusiones

- En esta misma línea, para que una muestra tomada a partir del PREP sea representativa, al menos, tendría que cumplir con:
 - Representatividad proporcional a nivel entidad federativa
 - Representatividad proporcional a nivel tipo de casilla
- Si bien es cierto que satisfacer las características anteriores es necesario, esto no es suficiente para poder llegar a un resultado concluyente, obligando a un análisis sobre las dispersiones de las medias estimadas
- La existencia de más o menos cruces entre las preferencias por el PAN y la coalición a lo largo de la captura del PREP no guarda relación con el resultado final o con la certidumbre o no del mismo

Conclusiones

- Ante la existencia de valores acotados, en la cantidad de votos por casilla, la ley de Benford no puede ser aplicada para detectar anomalías en el proceso de votación
 - Cada casilla tenía como máximo 750 boletas, más las de los funcionarios y representantes
 - Si al dato anterior, se le agrega un porcentaje determinado de abstencionismo, la cota superior incluso es menor a los 700 votos por casilla, en algunas casillas

Autores

- El presente análisis es un resultado del trabajo y la colaboración de:
 - Eduardo Ragasol
 - Carmen Cárdenas
 - Leonardo Muedano
 - Enrique Espinosa de los Monteros





Anexos





Anexo 1

Características de los Diseños



Características de los Diseño

- Diseño: Muestreo aleatorio simple
- Unidad de muestreo: Actas Contabilizadas por el PREP 2006
- Estimador:
 - La estimación del porcentaje nacional de votos válidos a favor del candidato presidencial x , será el cociente r_x donde:

$$r_x = \frac{\sum_{i=1}^n V_{xi}}{\sum_{i=1}^n V_i}$$

V_{xi} es el número de votos válidos emitidos en favor del candidato x en el acta i .
 V_i es el número de votos presidenciales totales emitidos en el acta i .
 n es el tamaño de muestra, en términos de actas y que en este caso toma los valores 1,000; 2,000 y 5,000

Características de los Diseños

- Diseño: Estratificado
- Unidad de muestreo: Actas Contabilizadas por el PREP 2006
- Estratificación:
 - De acuerdo a la entidad federativa
 - Al tipo de casilla (urbana y no urbana)
 - > Utilizando la estratificación descrita se crean 64 estratos
- Tamaño de muestra: 1,000 con asignación proporcional al peso del estrato, en términos del número de actas contabilizadas

Estimador para el Diseño Estratificado

- Estimador:
 - La estimación del porcentaje nacional de votos válidos a favor del candidato presidencial x , será el cociente r_x donde:

$$r_x = \sum_{h=1}^{64} W_h \frac{\sum_{i=1}^{n_h} V_{hxi}}{\sum_{i=1}^{n_h} V_{hi}}$$

V_{hxi} es el número de votos válidos emitidos en favor del candidato x en la casilla i que pertenece al estrato h .

V_{hi} es el número de votos presidenciales totales emitidos en la casilla i que pertenece al estrato h .

n_h es el tamaño de la muestra en el estrato h .

W_h es el peso del estrato h

$$W_h = \frac{\text{Actas Contabilizadas en el estrato } h}{\text{Actas Contabilizadas Totales}}$$

Donde: $\sum_{h=1}^{64} W_h = 1$

[Volver al análisis](#)



Anexo 2

Método Bootstrap



Método Bootstrap

- **Origen:**

- El método **bootstrap** aunque ya usado anteriormente, fue descrito de forma sistemática por Efron en 1979. El nombre alude al cordón de los zapatos, recordando la imagen de alguien intentando salir del barro tirando del cordón de sus propios zapatos

- **¿En que consiste?**

- Consiste en generar un gran número de muestras de tamaño n efectuando un muestreo con reemplazo de esos valores.
- En esa muestra calculamos el valor del parámetro que estamos estimando. Y así repetimos el proceso un gran número de veces, con lo que obtenemos una distribución de valores para el parámetro en la que podemos calcular su dispersión (análogo del error estándar) y determinar unos límites de confianza utilizando esa distribución.
- Es como si metiésemos los valores en una urna, extraemos una papeleta, anotamos el resultado y volvemos a colocarlo en la urna, y así hasta obtener N valores.
- El método bootstrap se considera como un caso particular del **Método de Monte Carlo**

[Volver al análisis](#)



Anexo 3

Pruebas de Bondad de Ajuste



¿En qué consiste la prueba de Bondad de Ajuste?

- A menudo se desea saber si las frecuencias observadas de ciertos sucesos difieren significativamente de las frecuencias esperadas para los mismos, es decir, si los conjuntos de ambas frecuencias se parecen o no
- Definición de X^2 (Chi-cuadrada)
 - Una medida de la discrepancia existente entre las frecuencias observadas y esperadas es suministrada por el estadístico X^2 , dado por:

$$X^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \dots + \frac{(o_k - e_k)^2}{e_k} = \sum_{j=1}^k \frac{(o_j - e_j)^2}{e_j}$$

– Donde:

> o_j son las j frecuencias observadas

> e_j son las j frecuencias teóricas o esperadas

¿En qué consiste la prueba de Bondad de Ajuste?

- Si $X^2 = 0$, las frecuencias observadas y esperadas concuerdan exactamente, mientras que si $X^2 > 0$, entonces no coinciden exactamente. A valores mayores de X^2 , mayores son las discrepancias entre las frecuencias observadas y esperadas.
- En la práctica, las frecuencias esperadas se calculan de acuerdo con la hipótesis H_0 . Si bajo esta hipótesis el valor calculado de X^2 dado es mayor que algún valor crítico, se deduce que las frecuencias observadas difieren *significativamente* de las esperadas y se rechaza H_0 al nivel de significación correspondiente. En caso contrario, no se rechazará. Este procedimiento se llama *ensayo* o *prueba de chi-cuadrado* de la hipótesis.

¿Cuál es la aplicación práctica de la prueba?

- La aplicación de la prueba de bondad de ajuste permite arrojar mayor claridad a las siguientes preguntas:
 - ¿Cómo saber si la muestra de casillas generada en cada corte del PREP, concuerda con los resultados teóricos?
 - ¿Cómo saber si estas muestras son representativas?
- Las pruebas que se desarrollan posteriormente mantienen las siguientes hipótesis:
 - H_0 : Las proporciones observadas concuerdan con las esperadas
 - H_1 : Las proporciones observadas no concuerdan con las esperadas
 - Los valores críticos para una prueba con nivel de significancia del 99% son:
 - > 52.19 para una X^2 con 31 grados de libertad (entidades)
 - > 6.63 para una X^2 con 1 grado de libertad (tipo de casilla)

¿Cuál es la aplicación práctica de la prueba?

- Específicamente para este análisis, se realizarán dos pruebas independientes, para comprobar si existe discrepancia entre las proporciones observadas y las esperadas

– Prueba 1. Se aplica a las proporciones esperadas para las 32 entidades federativas. X^2 está dado por:

$$X^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \dots + \frac{(o_{32} - e_{32})^2}{e_{32}} = \sum_{j=1}^{32} \frac{(o_j - e_j)^2}{e_j}$$

– Prueba 2. Se aplica a las proporciones esperadas por tipo de casilla (urbana y no urbana). X^2 está dado por:

$$X^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} = \sum_{j=1}^2 \frac{(o_j - e_j)^2}{e_j}$$

[Volver al análisis](#)



Anexo 4

Sustento del Análisis a los Cortes del PREP



Limitantes de las Muestras del PREP Tomadas en Distintos Cortes de Tiempo

- El Programa de Resultados Electorales Preliminares (PREP) ha sido un mecanismo institucional que permite dar a conocer el día de la Jornada Electoral los resultados preliminares de las elecciones federales, recabando los votos de las casillas asentados en las actas de escrutinio y cómputo, los cuales son difundidos de manera acumulada en diferentes cortes del tiempo, conforme van llegando y se van procesando en los centros de acopio y transmisión.
 - Las casillas que llegan en la primera hora a partir de que inicia la transmisión conforman una muestra dado que son una porción del total de casillas. Las casillas que se acumulan hacia la segunda hora, también son una muestra y así sucesivamente
- Sin embargo, estas muestras no cumplen con la definición de una muestra probabilística porque se desconoce la probabilidad con la llegará cada casilla en los diferentes cortes del PREP. Lo anterior supone que las muestras conformadas en los distintos cortes del tiempo, no concuerdan exactamente con los resultados teóricos esperados, según las reglas de probabilidad. Por ejemplo, si el universo de casillas tiene un 75% de urbanas y un 25% de no urbanas, el resultado esperado de una muestra probabilística de tamaño 100 es que 75 sean urbanas y 25 sean no urbanas.

Muestras PREP vs. Muestras Probabilísticas

- Ya se comentó que el conjunto de casillas cuyas actas llegan reportando un número de votos para cada partido o coalición a los largo del PREP, no cumplen con la definición de una muestra probabilística
- Por otro lado, la muestra especificada para un conteo rápido con diseño de muestreo probabilístico de esquema estratificado (aprovechando el conocimiento que se tiene de la población que se analiza) se espera que sí concuerde y reproduzca las características intrínsecas de las casillas y su votantes, sobre todo aquéllas características que se asocian a la preferencia de candidatos
- Por la experiencia de ACNielsen en conteos rápidos de elecciones presidenciales y el conocimiento que nos han aportado las encuestas de intención de voto, reconocemos **dos** características variables que se asocian a la preferencia del votante:
 - La entidad federativa donde está registrado y
 - El tamaño de localidad donde reside

[Volver al análisis](#)